# Improved approximate confidence intervals for censored data

by

Shuen-Lin Jeng

A dissertation submitted to the graduate faculty
in partial fulfillment of the requirements for the degree of
DOCTOR OF PHILOSOPHY

Major: Statistics
Major Professor: William Q. Meeker

Iowa State University
Ames, Iowa
1998

UMI Number: 9841053

Graduate College
Iowa State University

This is to certify that the Doctrol dissertation of
Shuen-Lin Jeng
has met the dissertation requirements of Iowa State University

Signature was redacted for privacy.
Major Professor

Signature was redacted for privacy.
For the Major Program

Signature was redacted for privacy.
For the Graduate College

# TABLE OF CONTENTS

# LIST OF TABLES

# LIST OF FIGURES

# ACKNOWLEDGMENTS

Studying in the Iowa State University has been a great experience. Here, the difference between weekend and weekdays is not clear to me. There is a man when ever I need to talk to I can reach him in his office or by email about 18 hours a day, 7 days a week, 12 months a year, for 5 years. He speaks to students loudly and clearly and listens to them carefully. Dr. Meeker is a mentor to me. He emphases these: theory and method, proof and intuition, science and praxis. Some of his characteristics are: assertive but patient, smart thinking and hard working, tough but soft. I am so lucky to have him to be my major professor. It is a pleasure for me to learn from him. And there is always more I could learn from him.

I would like to thank Necip Doganaksoy, Luis A. Escobar, C. Joseph Lu and Wayne Nelson for providing many useful suggestions on the first part of my dissertation.

Many thanks are due to Dr. Amemiya, Dr. Athreya, Dr. D. Cook, Dr. H. T. David, Dr. S. N. Lahiri, and Dr. B. Thompson for their valuable help and suggestions.

I would like to show my appreciation to those teachers and graduate students in Snedecor Hall who tolerate my never ending computer jobs on the Vincent workstations. Computing for this research was partly run on the DEC Alpha Farm of workstations at the Iowa State University Computation Center. Some of the simulations were done with equipment in the Statistics Department purchased with funds provided by an NSF SCREMS grant award DMS 9707740 to Iowa State University.

This research is partly supported by the Federal Aviation Administration Technical Center, Atlantic City, New Jersey, under Grant number 94G048.

My parents support me all the time. Especially my mother, Mrs. Yu-Chuei Huang. She always gives me wise advice. I would like to express my deepest gratitude to my sister and brother who care about me so much. Carrying their love I stand up and keep going forward.

# 1 GENERAL INTRODUCTION

In this research, we explore the censored data problem from parametric models. The following sections introduce the motivations of our research and address the directions of our approach.

## 1.1 Accurate Methods for Type I Censored Data

Due to time constraints in life testing, Type I censored data commonly arise from life tests. To make inference on parameters and quantiles of the life distribution, accurate confidence intervals (CIs) are needed. For Type II censored data (or uncensored data) from location-scale distributions (or log-location-scale distributions), Lawless (1982, page 147) presents pivotal quantities that can be used to obtain exact CIs for distribution parameters and quantiles analytically or through simulation. For Type I censoring (more common in practice), however, neither pivotal quantities nor other exact confidence interval methods in general exist.

Today, normal-approximation intervals are used most commonly in commercial software. These methods, however, may not have coverage probability close to nominal values for small to moderate number of failures, especially for one-sided confidence bounds. Methods for finding better approximate CIs is an important practical issue. Many papers in the literature discuss the coverage probability of two-sided CIs. Most methods do not perform equally well when one-sided confidence bounds (CBs) are concerned, even though most practical problems are one-sided (e.g., the cost of an error on one side is generally much different from the cost on the other side). We evaluate CI methods in order to find those that have high accuracy for both one-sided CBs and two-sided CIs and present those evaluations for both heavily censored and small sample cases.

We show some special effects of Type I censoring. With Type I censoring, unlike the complete data or Type II censoring case, the distributions of MLEs have a discrete component. Also the pivotal-like statistics have distributions that depend on the proportion failing. It is for these reasons that some bootstrap methods behave poorly in constructing confidence intervals for the $p$ quantile when $p$ is close to the proportion failing and the expected number of failures is small.

## 1.2 Performance of Bootstrap Likelihood Ratio Statistics

The asymptotic distributions of likelihood ratio statistics had been studied for decades. Most previous work has focused on the situations in which the underling distribution is continuous (especially parametric families) or discrete (e.g. empirical distributions). With time censored data, the distribution of a likelihood ratio statistics is a mixture of continuous and discrete parts. Jensen (1993) derives an Edgeworth expansion of log likelihood ratio (LLR) statistics for such data. For finding one-sided confidence intervals, the signed square root of log likelihood ratio (SRLLR) statistic is commonly used. This likelihood ratio statistic usually provides more accurate approximate inferences than the more commonly used studentized maximum likelihood estimators (see for example, Doganaksoy and Schmee (1993) and the first part of our study). However, generally the SRLLR statistics is approximated by the standard normal distribution only to first order $[O(1/n^{1/2})]$, even for complete continuous data (Barndorff-Nelsen 1994).

The bootstrap is a general procedure of resampling to find an approximate sampling distribution. We extend the results from Jensen (1993) and show that, under some regularity conditions, the distribution of the SRLLR statistics can be approximated up to second order accuracy $[O(1/n)]$ by using the bootstrap procedure.

## 1.3 Simultaneous Confidence Bands

In life testing and reliability studies, the primary problem of interest is often to estimate an unknown cumulative distribution function (cdf). For example, sample units are put on a life test. The purpose might be to estimate the proportion failing over a range of time points. Another example is the need to quantify nondestructive evaluation (NDE) capability. NDE methods are often used to detect a range of subsurface flaws before processing expensive materials. We want to know the detection ability for different flaw sizes. These problems can be formulated as one where an unknown cdf is to be estimated. We will use the more familiar failure time language in our general discussion.

Confidence intervals quantify the uncertainty of estimation. For example, pointwise confidence intervals with a specific confidence level can be computed for the cdf at particular times. When the interest is on the cdf for a range of times, the combination of these pointwise confidence intervals will not provide a simultaneous confidence band with same confidence level. Typically, for a given confidence level, a simultaneous confidence band would be wider than the joint set of pointwise confidence intervals. This is because we use the same amount of information from the data to do the inference for a specific point of interest as we use for inference on an infinite number of points.

Unlike pointwise confidence intervals, one cannot combine two $100(1-\alpha/2)\%$ one-sided simultaneous confidence bands to get a $100(1-\alpha)\%$ two-sided simultaneous confidence band. Different procedures are needed for one-sided and two-sided cases. Using the Wald statistics with Fisher information, Cheng and Iles (1983, 1988) provide a general procedure which can be applied to some continuous distribution that depends on a set of unknown parameters when data is complete. Censoring often arises in life data collection. Some theoretical results for complete data do not hold for censored data. Especially for Type I censoring, the Wald and likelihood ratio statistics no longer have the pivotal property (pivotal statistics have distributions that do not depend on unknown parameters) in location-scale models. The bootstrap method, however, provide a more accurate approximate distribution when the exact distributional form is not available. In the second part of our research we show that the bootstrap likelihood ratio statistics are generally second order accurate for complete and censored data. In the simulation study, we show that the bootstradissertationp Wald statistics with local information provide a confidence region with a coverage probability that appears to be as accurate as or more accurate than the bootstrap likelihood ratio statistics, even when the expected number of failures is small.

## 1.4 Dissertation Organization

The main body of this dissertation contains three papers that correspond to the problems raised in the previous sections. The first paper presented in Chapter 2 searches for accurate methods for Type I censored data. The second paper shown in Chapter 3 investigates the asymptotic performance of Bootstrap likelihood ratio statistics. The third paper in Chapter 4 explores the construction of simultaneous confidence bands. Those results presented in Chapter 3 and 4 can be applied to both complete and censored data. Chapter 5 gives the conclusion of this research.

## References

Barndorff-Nielsen, O. E., and Cox, D. R. (1994). *Inference and Asymptotics*. London: Chapman & Hall

Cheng, R. C. H. and Iles, T. C. (1983). "Confidence bands for cumulative distribution functions of continuous random variables," *Technometrics*, **25**, 77-86.

Cheng, R. C. H. and Iles, T. C. (1988). "One-sided confidence bands for cumulative distribution functions" *Technometrics*, **30**, 155-159.

Doganaksoy, N., and Schmee, J. (1993). "Comparisons of approximate confidence intervals for distributions used in life-data analysis," *Technometrics*, **35**, 175-184.

Jensen, J. L. (1993). "A historical sketch and some new results on the improved log likelihood ratio statistic," *Scandinavian Journal of Statistics, Theory and Applications*, **20**, 1-15.

Lawless, J. F. (1983). "Statistical methods in reliability," *Technometrics*, **25**, 305-316.

# 2 COMPARISONS OF WEIBULL DISTRIBUTION APPROXIMATE CONFIDENCE INTERVALS PROCEDURES FOR TYPE I CENSORED DATA

A paper submitted to the Technometics

Shuen-Lin Jeng and William Q. Meeker

## Abstract

This paper compares different procedures to compute confidence intervals for parameters and quantiles of the Weibull distribution for Type I censored data from life test experiments. The methods can be classified into three groups. The first group contains methods based on the commonly-used normal approximation for the distribution of (possibly transformed) studentized maximum likelihood estimators. The second group contains methods based on the likelihood ratio statistic and its modifications. The methods in the third group use a parametric bootstrap approach, including the use of bootstrap-type simulation to calibrate the procedures in the first two groups. All of these procedures are justified on the basis of large-sample asymptotic theory. We use Monte Carlo simulation to investigate the finite sample properties of these procedures. Our results show that the coverage probability of one-sided confidence bounds is much worse than those of two-sided confidence intervals calculated from methods in the first and second group. Usual normal-approximation methods are crude unless the expected number of failures is large (> 50 or 100). The likelihood ratio methods work much better and provide an adequate procedure down to 30 or 20 failures. The second-order bootstrap procedures do not perform equally well in small samples. By using bootstrap methods with caution, the coverage probability is close to nominal for expected number of failures down to 15 or less and even down to 10 or less for lightly censored cases (proportion failing > 50%). Exceptional cases, which are due to problems caused by the

Type I censoring, are noted.

**Keywords**: Bartlett correction, bias-corrected accelerated bootstrap, bootstrap-$t$, life data, likelihood ratio, ML estimator, parametric bootstrap, Type I censoring.

## 2.1 Introduction

### 2.1.1 Objectives

Due to time constraints in life testing, Type I censored data commonly arise from life tests. To make inference on parameters and quantiles of the life distribution, accurate confidence intervals (CIs) are needed. For Type II censored data (or uncensored data) from location-scale distributions (or log-location-scale distributions), Lawless (1982, page 147) presents pivotal quantities that can be used to obtain exact CIs for distribution parameters and quantiles analytically or through simulation. For Type I censoring (more common in practice), however, neither pivotal nor other exact confidence interval methods in general exist.

Today, normal-approximation intervals are used most commonly in commercial software. These methods, however, may not have coverage probability close to nominal values for small to moderate number of failures, especially for one-sided confidence bounds. Methods for finding better approximate CIs is an important practical issue. Many papers in the literature discuss the coverage probability of two-sided CIs. Most methods do not perform equally well when one-sided confidence bounds (CBs) are concerned, even though most practical problems are one-sided (e.g., cost of an error on one side is generally much different from the cost on the other side). We evaluate CI methods in order to find those that have high accuracy for both one-sided CBs and two-sided CIs and present those evaluations for both heavily censored and small sample cases.

We show some special effects of Type I censoring. With Type I censoring, unlike the complete data or Type II censoring case, the distributions of MLEs have a discrete component. Also the pivotal-like statistics have distributions that depend on the proportion failing. It is for these reasons that some bootstrap methods behave poorly in constructing confidence intervals for the $p$ quantile when $p$ is close to the proportion failing and the expected number of failures is small.

### 2.1.2 Related Work

For one-parameter distributions, exact confidence bound methods exist for the parameter or monotone functions of the parameter like distribution quantiles or failure probability (e.g., Mood, Graybill and Bose 1974, Section VIII.4 and Casella and Berger 1990, Section 9.2). When there are nuisance parameters, the situation is more complicated. For location-scale distributions, exact CIs can be obtained for parameters and some functions of the parameters based on complete or Type II censored data. For Type I censoring, using a model with one or more nuisance parameters, there are no known exact methods. Under usual regularity conditions, the large-sample approximate methods described in Section 2 work generally for distributions with two or more parameters.

In application, CIs based on normal-approximation theory (NORM method) of the ML estimator are popular. They are easy to calculate and the method has been implemented in most commercial software packages. Proper transformation of the ML estimator (TNORM method) can improve the approximation to the normal distribution. For example, statistics transformed to have a range over whole real line may perform closer to normal than those with finite boundaries.

Piegorsch (1987) explored the likelihood based interval for two-parameter exponential samples with Type I censoring. For the inference on the scale parameter, the coverage probabilities for two-sided CI becomes adequate when sample size reaches 25. Ostrouchov and Meeker (1988) showed that CIs based on inverting log likelihood ratio (LLR) tests provide better a approximation than TNORM CIs with interval censored data and Type I censoring for the Weibull and lognormal distributions. Vander Wiel and Meeker (1990) showed that for Type I censored Weibull data from case in accelerated life tests, LLR based CIs are better than those from the TNORM method.

Doganaksoy and Schmee (1993) compared four methods for Type I censored data from Weibull and lognormal distributions. They are NORM, LLR, the standardized LLR, and the LLR with Bartlett correction (LLRBART). They found that LLR-based methods perform much better than NORM intervals. With complete or moderately censored data, the standardized LLR considerably improves the approximation especially for small samples (down to 10 expected failures.) Doganaksoy (1995) reviewed likelihood ratio confidence intervals for reliability and life-data analysis applications. He notes that the LLRBART CIs have been used very little in applications due to the computational difficulties of implementation.

Recent research indicates that the bootstrap is a very powerful method for computing accurate approximate confidence intervals. Hall (1987, 1992), Efron and Tibshirani (1993), Shao and Tu (1995) describe bootstrap theory and methods in detail. The parametric bootstrap method mimics the distribution of statistics by simulation or re-sampling.

Robinson (1983) applied the bootstrap method to location and scale distributions. The statistics used for constructing confidence intervals are pivotal quantities in the case of complete or Type II censored data. He used the method to find CIs for multiple time-censored progressive data and used simulation to evaluate coverage probabilities.

The parametric bootstrap-$t$ (PBT) is second-order accurate under smoothness conditions (Efron 1982). The percentile method (Efron 1981) is very easy to implement but usually is only first-order accurate for one-sided CBs. The bias-corrected method (BC, Efron 1982) generally has better performance than the percentile method. The bias-corrected accelerated method (BCA, Efron 1987) provides an alternative, more accurate, method to construct CIs that usually improves the performance of percentile and BC method in complete samples.

The signed-root log-likelihood ratio (SRLLR) statistic has an approximate normal distribution in large samples (Barndorff-Nielsen and Cox 1994, page 101). Modified SRLLR method (Barndorff-Nielsen 1986, 1991) is third-order accurate in complete samples but needs much more efforts to get the modification term. Using bootstrap simulation to obtain the sampling distribution of the SRLLR statistic (PBSRLLR), instead of using the large-sample approximate distribution (normal), improves the procedure's coverage probabilities, especially for one-sided CBs. PBSRLLR method is different from the one that uses bootstrap simulation procedure to approximate the distribution of LLR statistic (PBLLR, see Appendix A.2) and has accuracy has accuracy better than the PBLLR method for one-sided cases.

### 2.1.3 Overview

The remainder of this paper is organized as follows. Section 2 describes the model and the estimation method. Section 3 provides details of the methods for finding approximate CIs. Section 4 describes the design of the simulation experiment. Section 5 presents the general results from the simulation experiment. Section 6 contains conclusions from the experiment and suggestions for use in applications. Section 7 discusses some special effects of Type I censoring that lead to poor performance of some simulation-based CI/CB procedures. Discussion and directions for future research are given in Section 8.

## 2.2 Model and Estimation

### 2.2.1 Model

If $T$ has a Weibull distribution, then $Y = \log(T)$ has a smallest extreme value (SEV) distribution with density

$$f_Y(y) = \frac{1}{\sigma}\exp\left[\frac{y-\mu}{\sigma} - \exp\left(\frac{y-\mu}{\sigma}\right)\right],$$

and cdf

$$F_Y(y) = 1 - \exp\left[-\exp\left(\frac{y-\mu}{\sigma}\right)\right],$$
$$-\infty < y < \infty, -\infty < \mu < \infty, \sigma > 0,$$

where $\mu$ and $\sigma$ are location and scale parameters. The $q$ quantile of the SEV distribution is $y_q = F_Y^{-1}(q) = \mu + c_q\sigma$, where $c_q = \log[-\log(1-q)]$ is the $q$ quantile of the standardized ($\mu = 0$, and $\sigma = 1$) SEV distribution. Define $\alpha = \exp(\mu)$ and $\beta = 1/\sigma$ as Weibull parameters.

### 2.2.2 ML Estimation

We use $\widehat{\mu}$ and $\widehat{\sigma}$ to denote the ML estimators of the SEV parameters. Because of the invariance property of ML estimators, $\widehat{y}_q = \widehat{\mu} + c_q\widehat{\sigma}$ is the ML estimator of the $q$ quantile of the SEV distribution. Also the ML estimators of the Weibull parameters are $\widehat{\alpha} = \exp(\widehat{\mu})$ and $\widehat{\beta} = 1/\widehat{\sigma}$. The ML estimator of the $q$ Weibull quantile is $\widehat{t}_q = \exp(\widehat{y}_q)$. Also $\widehat{y}_q = \widehat{\mu} + c_q\widehat{\sigma}$ is the ML estimator of the $q$ quantile of the SEV distribution. More generally the ML estimator of a function $\boldsymbol{g}(\mu, \sigma)$ is $\widehat{\boldsymbol{g}} = \boldsymbol{g}(\widehat{\mu}, \widehat{\sigma})$. For any function of interest, it is possible to re-parameterize by defining a one-to-one transformation, $\boldsymbol{g}(\mu, \sigma) = (g_1(\mu, \sigma), g_2(\mu, \sigma)) = \boldsymbol{\theta}$, that contains the function of interest among its elements. For example $g_1(\mu, \sigma)$ could be a distribution quantile or failure probability. Then ML fitting can be carried out for this new parameterization in a manner that is the same as that described above for $(\mu, \sigma)$. This provides a procedure for obtaining ML estimates and likelihood confidence intervals for any scalar or vector function of $(\mu, \sigma)$. For more details see Lawless (1982, Chapter 4).

Let $\boldsymbol{\theta} = (\theta_1, \theta_2)$ be the unknown parameter vector, where $\theta_1$ is the parameter of interest and $\theta_2$ is a nuisance parameter. Typically $\boldsymbol{\theta}$ could be $(\mu, \sigma)$ or $(t_q, \sigma)$. $L(\boldsymbol{\theta})$ is the likelihood and let $t_c$ denote the specified censoring time. Let $t_1, \ldots, t_n$ be $n$ observations

Table 2.1 Abbreviations for CB/CI methods

| | |
|---|---|
| NORM | Normal-approximation |
| TNORM | Transformed normal-approximation |
| LLR | Log likelihood ratio |
| LLRBART | Log likelihood ratio Bartlett corrected |
| PBT | Parametric bootstrap-$t$ |
| PTBT | Parametric transformed bootstrap-$t$ |
| PBSRLLR | Parametric bootstrap signed square root LLR |
| PBP | Parametric bootstrap percentile |
| PBBCA | Parametric bootstrap bias-corrected accelerated |
| PBBC | Parametric bootstrap bias-corrected |

(e.g., failure or censoring times) from a life test. If the observations are independent, then the censored-data likelihood is

$$L(\boldsymbol{\theta}) = \prod_{i=1}^{n} [f_Y(\log(t_i); \boldsymbol{\theta})]^{\delta_i} [1 - F_Y(\log(t_c); \boldsymbol{\theta})]^{1-\delta_i},$$

where $\delta_i = 1$ if $t_i$ is a failure time and $\delta_i = 0$ if observation $i$ is censored at $t_c$.

## 2.3 Confidence Bound/Interval Methods

This section describes the different CI/CB procedures that we study in this paper. For more details, see the given references. Table 2.1 shows the abbreviation for each method. Let $C_{n;1-\alpha}$ denote an approximate CI for $\theta_1$ with nominal coverage probability $1-\alpha$, where $n$ is the sample size. The procedure for obtaining $C_{n;1-\alpha}$ is said to be $k$th order accurate if $\Pr(\theta_1 \in C_{n;1-\alpha}) = 1 - \alpha + O(n^{-k/2})$. If there is no $O(\cdot)$ term in the equation, we say that the procedure for $C_{n;1-\alpha}$ is "exact."

### 2.3.1 Normal-Approximation Methods

**Normal-approximation method (NORM).** Suppose $\widehat{\boldsymbol{\theta}}$ is the ML estimator of the parameter vector $\boldsymbol{\theta}$. Under the usual regularity conditions, $\widehat{\boldsymbol{\theta}}$ is asymptoticly normal and efficient (Serfling 1980, page 148). Let $I_{\boldsymbol{\theta}}$ denote the Fisher information matrix and $n[\widehat{\mathrm{se}}(\widehat{\theta}_1)]^2$ be an estimator that converges to $I_{\boldsymbol{\theta}}^{(1,1)}$ in probability when $n$ increases to $\infty$, where $I_{\boldsymbol{\theta}}^{(1,1)}$ is the $(1,1)$ term of the inverse of $I_{\boldsymbol{\theta}}$. Then the distribution of $(\widehat{\theta}_1 - \theta_1)/\widehat{\mathrm{se}}(\widehat{\theta}_1)$

is approximately $N(0,1)$ in large samples. A normal-approximation $100(1-\alpha)\%$ confidence interval can be obtained from $\widehat{\theta}_1 \pm z_{(1-\alpha/2)}\widehat{\mathrm{se}}(\widehat{\theta}_1)$, where $z_{(1-\alpha/2)}$ is the $N(0,1)$ distribution $1-\alpha/2$ quantile. In this paper $n[\widehat{\mathrm{se}}(\widehat{\theta}_1)]^2$ is obtained from the inverse of the local estimate of the $I_{\boldsymbol{\theta}}$ (e.g., Nelson 1982, page 377).

**Transformed normal-approximation method (TNORM).** When an ML estimator $\widehat{\theta}_1$ has its range on only part of the real line, a monotone continuously differentiable function $g(\widehat{\theta}_1)$ with range on the entire real line could have a better normal-approximation (Nelson, 1982, page 331). Let $g'(\theta_1)$ denote the derivative of $g(\theta_1)$ and let $n\{\widehat{\mathrm{se}}[g(\widehat{\theta}_1)]\}^2$ be an estimator that converges to $g'(\theta_1)I_{\boldsymbol{\theta}}^{(1,1)}g'(\theta_1)$ in probability. Using the delta method, $[g(\widehat{\theta}_1)-g(\theta_1)]/\widehat{\mathrm{se}}[g(\widehat{\theta}_1)] \mathrel{\dot\sim} N(0,1)$. Then an approximate confidence interval for $\theta_1$ can be obtained from $g^{-1}\{g(\widehat{\theta}_1) \pm z_{(1-\alpha/2)}\widehat{\mathrm{se}}[g(\widehat{\theta}_1)]\}$, where $z_{(1-\alpha/2)}$ is the $1-\alpha/2$ quantile of the $N(0,1)$ distribution. Typically $g$ could be the log function for a scale parameter or for positive quantile parameters and the logit or tanh function for a probability parameter. In this paper $n\{\widehat{\mathrm{se}}[g(\widehat{\theta}_1)]\}^2$ is taken to be $g'(\widehat{\theta}_1)\widehat{I}_{\widehat{\boldsymbol{\theta}}}^{(1,1)}g'(\widehat{\theta}_1)$, where $\widehat{I}_{\widehat{\boldsymbol{\theta}}}$is the local estimate of $I_{\boldsymbol{\theta}}$.

### 2.3.2 Likelihood Ratio Methods

**Log LR method (LLR).** The profile likelihood for $\theta_1$ is defined as

$$R(\theta_1) = \max_{\theta_2}\left[\frac{L(\theta_1,\theta_2)}{L(\widehat{\boldsymbol{\theta}})}\right]. \tag{2.1}$$

Let $W = -2\log R(\theta_1)$. From Serfling (1980, Section 4.4), the limiting distribution of $W$ is $\chi^2_1$. Thus an approximate $100(1-\alpha)\%$ confidence interval can be calculated from $\min\{W^{-1}(\chi^2_{(1-\alpha,1)})\}$ and $\max\{W^{-1}(\chi^2_{(1-\alpha,1)})\}$, where $W^{-1}[\cdot]$ is the inverse mapping and $\chi^2_{(1-\alpha,1)}$ is the $1-\alpha$ quantile of $\chi^2$ distribution with 1 degree of freedom.

**Log LR Bartlett corrected method (LLRBART).** Because the expectation of $W/E(W)$ is equal to the mean of the $\chi^2_1$ distribution, the distribution of $W/E(W)$ will be better approximated by the $\chi^2_1$ distribution (Bartlett 1937). In general one must substitute an estimate for $E(W)$ computed from one's data. For complicated problems (e.g., those involving censoring) it is necessary to estimate of $E(W)$ by using simulation. Then an approximate $100(1-\alpha)\%$ confidence interval can be obtained by using $\min\{W^{-1}[\chi^2_{(1-\alpha,1)}\widehat{E}(W)]\}$ and $\max\{W^{-1}[\chi^2_{(1-\alpha,1)}\widehat{E}(W)]\}$.

### 2.3.3 Parametric Bootstrap Methods

The following methods use the "bootstrap principle" or Monte Carlo evaluation of sampling distributions. Suppose a statistic $S$ is a function of random variables with a distribution that depends on the parameter $\boldsymbol{\theta}$. The parametric bootstrap version $S^*$ of $S$ is the same function but evaluated at data ("bootstrap sample") simulated using $\widehat{\boldsymbol{\theta}}$ instead of the unknown $\boldsymbol{\theta}$. Using $\widehat{\boldsymbol{\theta}}$ in place of the distribution parameters, the distribution of $S^*$ can be calculated analytically in simple situations, or by simulation in general.

**Parametric bootstrap-$t$ method (PBT).** (Efron 1982) Let $\widehat{\theta}_1$ be the ML estimator of $\theta_1$ and let $\widehat{\theta}_1^*$ be the ML estimator from bootstrap data. Also let $z_{\widehat{\theta}^*_{1(\alpha)}}$ be the $\alpha$ quantile of the distribution of $(\widehat{\theta}_1^* - \widehat{\theta}_1)/\widehat{\text{se}}^*(\widehat{\theta}_1)$, where $\widehat{\text{se}}^*(\widehat{\theta}_1)$ is the bootstrap version of $\widehat{\text{se}}(\widehat{\theta}_1)$. In this paper we choose $\widehat{\text{se}}(\widehat{\theta}_1)$ to be the same as in the NORM method. The approximate $100(1-\alpha)\%$ confidence limits are computed from $\widehat{\theta}_1 - z_{\widehat{\theta}^*_{1(1-\alpha/2)}}\widehat{\text{se}}(\widehat{\theta}_1)$ and $\widehat{\theta}_1 - z_{\widehat{\theta}^*_{1(\alpha/2)}}\widehat{\text{se}}(\widehat{\theta}_1)$.

**Parametric transformed bootstrap-$t$ method (PTBT).** Let g be a smooth monotone function generally chosen such that $g(\widehat{\theta}_1)$ has range on whole real line. Let $\widehat{\theta}_1$ be the ML estimator of $\theta_1$ and let $\widehat{\theta}_1^*$ be the bootstrap version ML estimator. Let $z_{g(\widehat{\theta}_1^*)_{(\alpha)}}$ be the $\alpha$ quantile of the distribution of $[g(\widehat{\theta}_1^*) - g(\widehat{\theta}_1)]/\widehat{\text{se}}^*[g(\widehat{\theta}_1)]$, where $\widehat{\text{se}}^*[g(\widehat{\theta}_1)]$ is the bootstrap version of $\widehat{\text{se}}[g(\widehat{\theta}_1)]$. In this paper we choose $\widehat{\text{se}}[g(\widehat{\theta}_1)]$ to be the same as in the TNORM method. An approximate $100(1-\alpha)\%$ confidence interval for $\theta_1$ can be computed from $g^{-1}\{g(\widehat{\theta}_1) - z_{g(\widehat{\theta}_1^*)_{(1-\alpha/2)}}\widehat{\text{se}}[g(\widehat{\theta}_1)]\}$ and $g^{-1}\{g(\widehat{\theta}_1) - z_{g(\widehat{\theta}_1^*)_{(\alpha/2)}}\widehat{\text{se}}[g(\widehat{\theta}_1)]\}$.

**Parametric bootstrap signed square root log LR method (PBSRLLR).** Let $V(\theta_1) = \text{sign}(\widehat{\theta}_1 - \theta_1)[-2\log R(\theta_1)]^{1/2}$ denote the signed square root of the log likelihood ratio statistic. In large samples, the distribution of $V(\theta_1)$ can be approximated by a normal distribution. Obtaining the distribution by simulation, however, captures the asymmetry of the distribution and hence provides a better approximation for finding confidence bounds for $\theta_1$. Suppose that $v_{\widehat{\theta}^*_{1(\alpha)}}$ is the $\alpha$ quantile of the bootstrap distribution of $V(\theta_1)$. Then an approximate $100(1-\alpha)\%$ confidence interval can be computed from $\min\{V^{-1}(v_{\widehat{\theta}^*_{1(\alpha/2)}}), V^{-1}(v_{\widehat{\theta}^*_{1(1-\alpha/2)}})\}$ and $\max\{V^{-1}(v_{\widehat{\theta}^*_{1(\alpha/2)}}), V^{-1}(v_{\widehat{\theta}^*_{1(1-\alpha/2)}})\}$.

**Parametric bootstrap percentile method (PBP).** (Efron 1981) Let $\widehat{\theta}_1$ be the ML estimator of $\theta_1$ and let $\widehat{\theta}_1^*$ be the bootstrap version of the ML estimator. Suppose $\widehat{\theta}^*_{1(\alpha)}$ is the $\alpha$ quantile of the distribution of $\widehat{\theta}_1^*$. Then an approximate $100(1-\alpha)\%$ confidence interval for $\theta_1$ can be computed from $\widehat{\theta}^*_{1(\alpha/2)}$ and $\widehat{\theta}^*_{1(1-\alpha/2)}$. If there exists an increasing transformation $v_n(\theta_1)$ and $\Psi$ a continuous, increasing and symmetric distribution such that

$$\Pr\{v_n(\widehat{\theta}_1) - v_n(\theta_1) \le x\} = \Psi(x)$$

holds, then the PBP CI procedure is exact. Otherwise the one-sided PBP CBs are only first order accurate. Note that the forms of $v_n$ and $\Psi$ need not be known to compute the interval.

**Parametric bootstrap bias-corrected accelerated method (PBBCA).** Based on concerns expressed on Schenker and Patwardhan (1985), Efron (1987) suggested an improved percentile bootstrap method that corrected for both bias and non-constant scale and named it BCA (bias-corrected and accelerated) method. Efron and Tibshirani (1993, Section 14.3) showed an easier way to obtain BCA confidence intervals. An approximate $100(1-\alpha)\%$ confidence interval is given by $(\widehat{\theta}^*_{1(\alpha_1)}, \widehat{\theta}^*_{1(\alpha_2)})$. Where $\widehat{\theta}^*_{1(\alpha)}$ is the $\alpha$ quantile of the distribution of $\widehat{\theta}^*_1$ and

$$\alpha_1 = \Phi\left(\widehat{z}_0 + \frac{\widehat{z}_0 + z_{\alpha/2}}{1 - \widehat{a}(\widehat{z}_0 + z_{\alpha/2})}\right), \quad \alpha_2 = \Phi\left(\widehat{z}_0 + \frac{\widehat{z}_0 + z_{1-\alpha/2}}{1 - \widehat{a}(\widehat{z}_0 + z_{1-\alpha/2})}\right),$$

$$\widehat{z}_0 = \Phi^{-1}\left(\frac{\#\{\widehat{\theta}^*_1(b) < \widehat{\theta}_1\}}{B}\right), \quad \widehat{a} = \frac{\sum_{i=1}^n \left(\widehat{\theta}_{1[\cdot]} - \widehat{\theta}_{1[i]}\right)^3}{6\left[\sum_{i=1}^n \left(\widehat{\theta}_{1[\cdot]} - \widehat{\theta}_{1[i]}\right)^2\right]^{3/2}}.$$

Usually $\Phi$ is taken to be the standard normal cdf. Here $\widehat{\theta}_{1[i]} = \widehat{\theta}_1(X_{[i]})$, $X_{[i]}$ is the original sample with the $i$th point $x_i$ deleted, $\widehat{\theta}_{1[\cdot]} = \sum_{i=1}^n \widehat{\theta}_{1[i]}/n$, $z_\alpha$ is the $\alpha$ quantile of normal distribution, $B$ is the number of the bootstrap samples, and $\widehat{\theta}^*_1(b), b = 1, \ldots, B$ are bootstrap versions of $\widehat{\theta}_1$.

If there is an increasing function $v_n$ (the exact form need not be known) such that

$$\Pr\left\{\frac{v_n(\widehat{\theta}_1) - v_n(\theta_1)}{1 + a v_n(\theta_1)} + z_0 \le x\right\} = \Phi(x),$$

then the BCA CI procedure is exact.

**Parametric bootstrap bias-corrected method (PBBC).** Suppose that there exists an increasing function $v_n$ (the exact form could be unknown), a cumulative distribution function $\Phi(x)$ (the exact form needs to be specified) and

$$\Pr\left\{v_n(\widehat{\theta}_1) - v_n(\theta_1) + z_0 \le x\right\} = \Phi(x).$$

Efron (1982) showed that the exact CI procedure for $\theta_1$ can be obtained. For most cases the form of $\Phi(x)$ is not available, and the standard normal cdf is suggested for $\Phi$. This is the special case of the PBBCA method and can be calculated as in the PBBCA method by putting $\hat{a} = 0$.

## 2.4 Simulation Experiment

This section describes our simulation experiment to compare the different confidence interval procedures.

### 2.4.1 Simulation Design

Our simulation experiment was designed to study the following factors:

- $p_f$: the expected proportion failing by the censoring time.
- $E(r) = np_f$: the expected number of failures before the censoring time.

We used 2000 Monte Carlo samples for each $p_f$ and $E(r)$ combination. The levels used were $p_f$ = .01, .05, .1, .3, .5, .7, .9, 1 and $E(r)$ = 3, 5, 7, 10, 15, 20, 30, 50 and 100. For each Monte Carlo sample we obtained the ML estimates of the scale parameter and the quantiles $y_q$, $q$ = .01, .05, .1, .3, .5, .632 and .9, where $\mu \cong y_{.632}$. The one-sided $100(1-\alpha)\%$ confidence bounds were calculated for $\alpha$ =.025 and .05. Hence the two-sided CIs, 90% and 95%, can be obtained by combining the upper and lower CBs. Without loss of generality, we sampled from an SEV distribution with $\mu = 0$ and $\sigma = 1$.

Because the number of failures before the time censoring $t_c$ is random, it is possible to have as few as $r = 0$ or 1 failures in the simulation, especially when $E(r)$ is small. With $r = 0$, ML estimates do not exist. With $r = 1$, LR intervals may not exist. Therefore, we calculate the results conditionally on the cases with $r > 1$, and report the observed nonzero proportions that resulted in $r \leq 1$.

### 2.4.2 Parameter Estimation and Computation Methods

ML estimates of $\mu$ and $\sigma$ were obtained by solving the simultaneous equations in Appendix A.1.1. For finding the confidence limits from LR methods, two equations from the log likelihood were used. The first one specifies the quantile to be estimated.

The second equation assures that the constrained ML estimator will be achieved for the nuisance parameter. See the Appendix A.1.2-A1..4 for further details.

The Fortran subroutines DNSQE and NNES from netlib (http://www.netlib.org) were used to solve the simultaneous nonlinear equations. The TNORM confidence limits were used as starting values. For the small expected number failing cases ($E(r) < 10$) and heavy censoring ($p_f < .2$), the start values were not always close enough to the solution of equations. Two methods were attempted to overcome the difficulties. First we switched from the Powell hybrid method to a line search method. If line search failed, as a last resort, we used different sets of start values obtained from a grid search and this method was always successful. The program was written in Fortran with calculations performed in double precision. The accuracy for the parameter estimates and the CI calculation was approximately 6 significant digits.

The computer time required for the simulation is an increasing function of the expected number of failures. Simulating bootstrap intervals is computationally intensive. For $E(r) = 3$ and $p_f = 1$ it takes about 10 seconds to calculate one Monte Carlo simulation trial for all CIs for different methods and parameters. For $E(r) = 100$ and $p_f = 1$ it takes approximately 80 seconds. We used 2000 replications in the simulation. Most of the simulations were done using DEC 3000 Model 900 Alpha workstations.

### 2.4.3 Coverage Probability Comparisons

Let $1 - \alpha$ be the nominal coverage probability (CP) of a CI, and let $1 - \widehat{\alpha}$ denote the corresponding Monte Carlo estimate. The standard error of $\widehat{\alpha}$ is approximately $\text{se}(1 - \widehat{\alpha}) = [\alpha(1 - \alpha)/n_s]^{1/2}$, where $n_s$ is the number of Monte Carlo simulation trials. For one-sided 95% CBs from 2000 simulations the standard error of CP estimation is $[.05(1 - .95)/2000]^{1/2} = .0049$. The Monte Carlo error is approximately $\pm 1\%$. We say the method is adequate if the CP is within $\pm 2\%$ error for 95% CBs and 90% CIs.

If the estimated actual coverage probability is greater than (less than) $1 - \alpha$ then the CI procedure is conservative (anti-conservative). We say that coverage probability is approximately symmetric when the difference of the CP of lower and upper CBs is approximately the same.

## 2.5 Results of Simulation Experiment

The results of the simulation study were summarized using different numerical and graphical methods. Here we present some of the most interesting and useful graphical

Table 2.2 Number of the cases where $r = 0$ or 1 in 2000 Monte Carlo simulations of the experiment. The expected numbers rounded to the nearest integer are shown inside parentheses.

| | | $p_f$ | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | .01 | .05 | .10 | .30 | .50 | .70 | .90 |
| $E(r)$ | 3 | 379(395) | 365(383) | 376(367) | 308(298) | 235(218) | 160(167) | 63( 55) |
| | 5 | 88( 79) | 72( 74) | 68( 67) | 59( 52) | 23( 21) | 11( 7) | 1( 0) |
| | 7 | 17( 14) | 16( 12) | 13( 10) | 3( 5) | 1( 1) | 0( 0) | 1( 0) |
| | 10 | 0( 0) | 3( 0) | 0( 0) | 0( 0) | 0( 0) | 0( 0) | 0( 0) |

displays. Table 2.2 shows the number of Monte Carlo simulations which that had only 0 or 1 failure. Those cases are excluded in calculating coverage probability. When $E(r) > 10$, there was no Monte Carlo simulation that had fewer than 2 failures.

### 2.5.1 One-sided CBs

Let UCB (LCB) denote an upper (lower) confidence bound. Figure 2.1 shows the coverage probability of the one-sided approximate 95% CBs for the parameter $\sigma$ from 10 methods for 5 cases of proportion failing. Figure 2.2 is the same type of graph as Figure 2.1 for the .1 quantile, $t_{.1}$, of Weibull distribution. The crossing of lines for some cases with $E(r) = 3$ and 5 in Figure 2.2 is due to dropping the simulation trials where $r = 0$ or 1. Figure 2.3 shows CPs when $p_f = .5$ for different quantiles. Figure 2.4 to Figure 2.7 present a closer comparison of CP for methods and parameters.

For the parameter $\sigma$, we have following results from Figure 2.1:

- For the NORM method both UCBs and LCBs have inaccurate CP for all $p_f$, even when $E(r) = 50$. UCBs are always anti-conservative and LCBs are always conservative. The TNORM method has CPs closer to the nominal ones for $E(r) \geq 30$ but the confidence bounds still have the same asymmetry as in NORM method.

- For the LLR method, as $E(r) < 20$, UCBs are anti-conservative and LCBs are conservative. For $E(r) \geq 20$ and $p_f \leq .5$, the approximation is adequate. The LLRBART method does not improve the CP relative to the LLR.

- The PBBCA, PBBC and PBP methods have CPs approximately equal to the nominal CPs for $E(r)$ greater than 20, 50, 100, respectively.

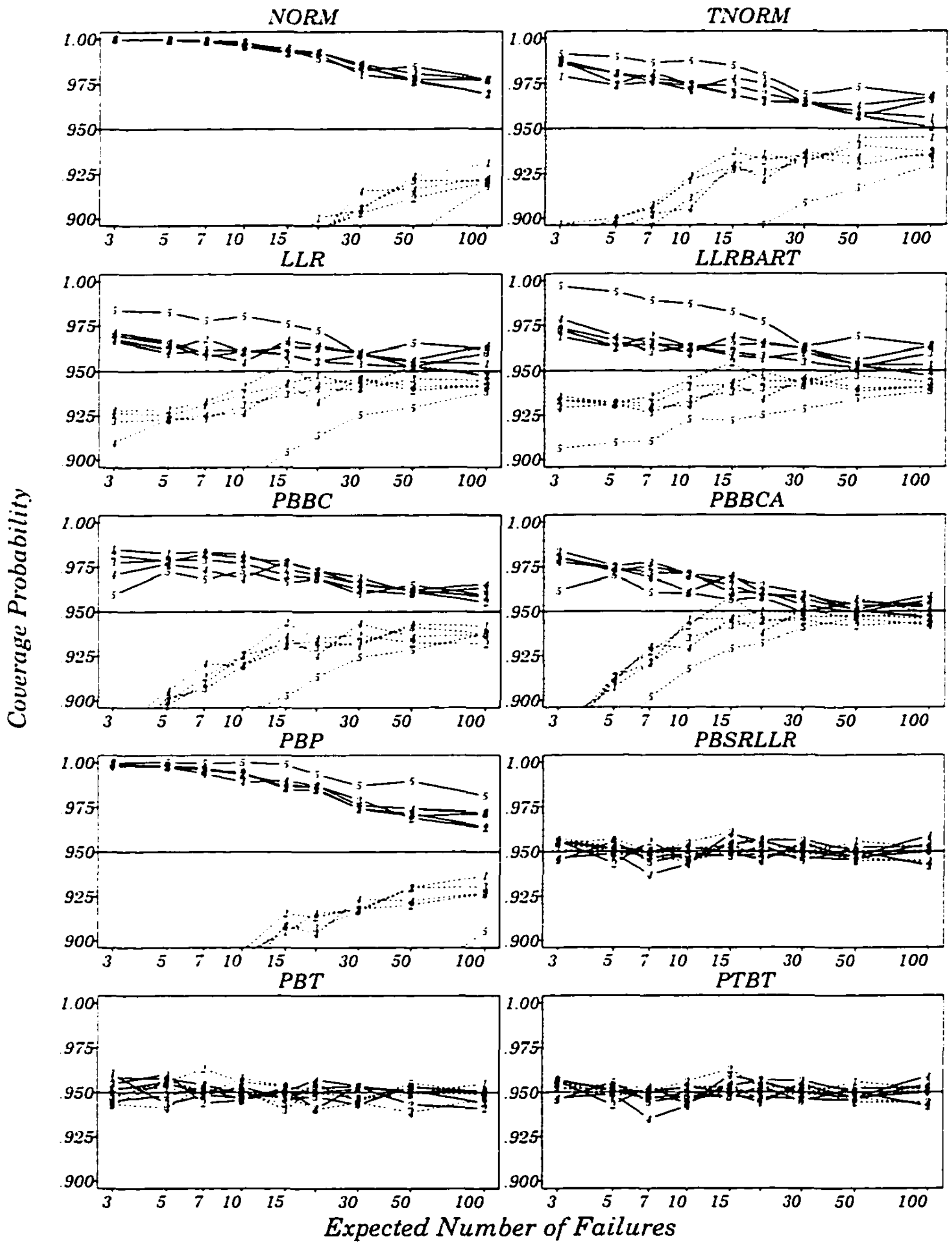



Figure 2.1 Coverage probability versus expected number of failures plot of one-sided approximate 95% CIs for parameter $\sigma$. The numbers (1, 2, 3, 4, 5) in the lines of each plot correspond to $p_f$'s (.01, .1, .3, .5, 1). Dotted and solid lines correspond to upper and lower bounds, respectively.

- The PBSRLLR, PBT and PTBT methods always provide excellent approximations even for $E(r) = 3$ case.

For estimating distribution quantiles, the situation is more complicated. In general, for quantiles $t_p$ (Figure 2.2 to Figure 2.7):

- For NORM, UCBs are anti-conservative and LCBs are conservative in most cases when $E(r) \geq 10$. The approximation of CP is crude (e.g., $1 - \widehat{\alpha} < .90$ for nominal $1 - \alpha = .95$) for some parameters even in the case $E(r) = 100$. NORM has better performance for quantiles $t_p$ for which $p < p_f$.

- TNORM is more accurate than NORM for $E(r) > 30$. The approximation of CP is still crude and depends on $p_f$. UCBs (LCBs) are conservative when $p < p_f$ ($p > p_f$) and are anti-conservative when $p > p_f$ ($p < p_f$) except that when $p$ from $t_p$ is close to $p_f$, both are conservative.

- The asymmetry for the LLR and LLRBART is similar to TNORM, but the CPs are closer to the nominal ones. That is, UCBs (LCBs) are conservative when $p < p_f$ ($p > p_f$) and are anti-conservative when $p > p_f$ ($p < p_f$). LLR provides good CP for $E(r) \geq 20$. LLRBART improves LLR only when $p_f \geq .7$.

- PBT has poor CP even for $E(r) = 100$. It is a little more accurate with no censoring than in the censored cases. Depending on the particular quantiles, PBT could be better or worse than NORM.

- When $p_f \geq .5$, the ordering with respect to CP accuracy, in descending order, is PBBCA, PBBC and PBP. Otherwise there is no strict order for these three methods. UCBs are always conservative and LCBs are anti-conservative. Generally PBBCA and PBBC have adequate CP for $E(r) > 20$. But for $p_f < .5$ and $p > p_f$, PBBC is better than PBBCA. When $p \approx p_f$, PBBC and PBBCA have lower CP than in other cases.

- For PBSRLLR and PTBT, UCBs and LCBs all provide excellent approximations when $p_f < p$ especially for heavily censored cases ($p_f < .1$). But when $p_f$ is close to $p$, both methods have lower CP for LCBs. The PBSRLLR method is better than the PTBT method and is adequate for $E(r) \geq 15$. When $p_f \geq .5$, the PBSRLLR method is adequate for $E(r) \geq 10$.

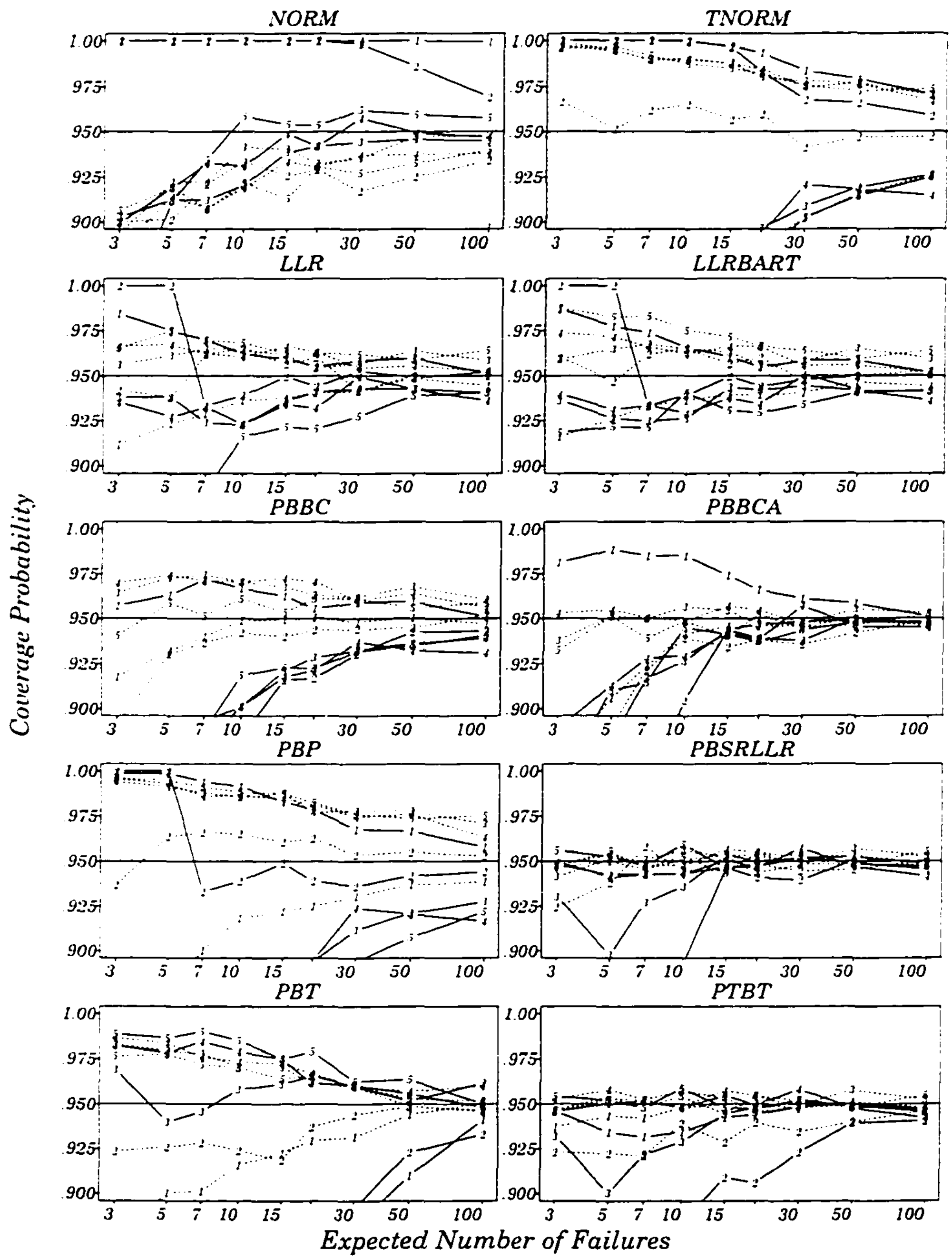



Figure 2.2 Coverage probability versus expected number of failures plot of one-sided approximate 95% CIs for parameter $t_{.1}$. The numbers (1, 2, 3, 4, 5) in the lines of each plot correspond to $p_f$'s (.01, .1, .3, .5, 1). Dotted and solid lines correspond to upper and lower bounds, respectively.

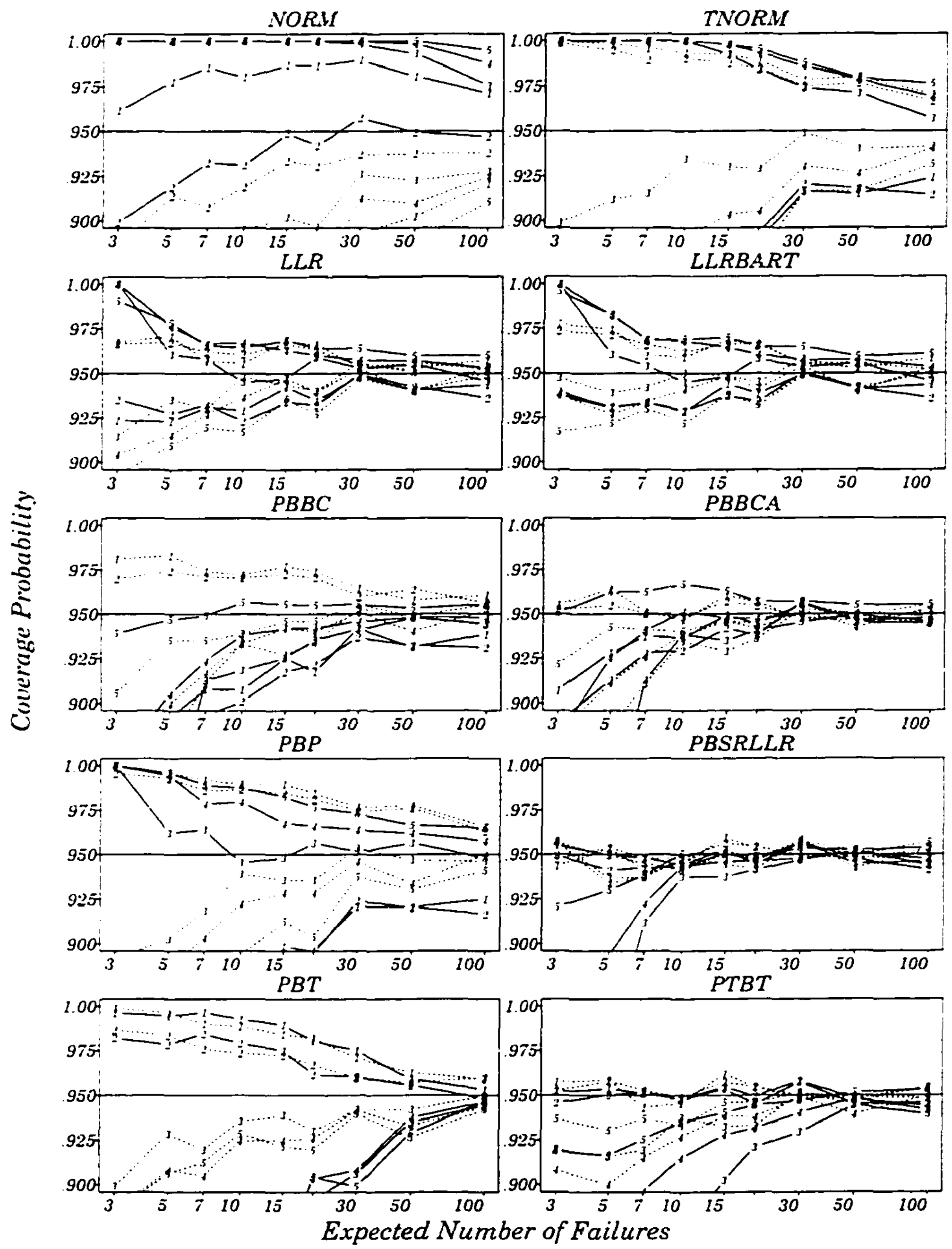



Figure 2.3 Coverage probability versus expected number of failures plot of one-sided approximate 95% CIs for proportion failing $p_f = .5$. The numbers (1, 2, 3, 4, 5) in the lines of each plot correspond to $t_p$'s, $p = (.01, .1, .5, .632, .9)$. Dotted and solid lines correspond to upper and lower bounds, respectively.

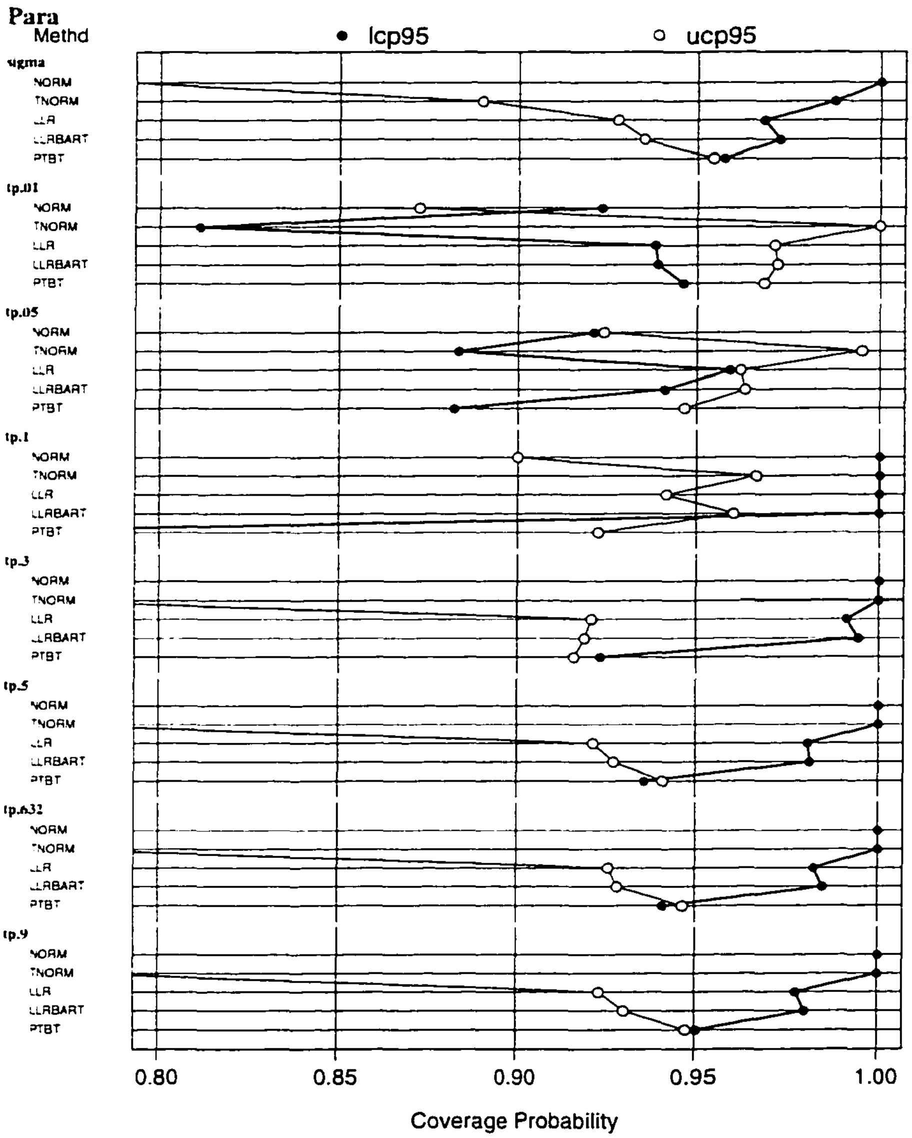



Figure 2.4 Coverage probability plot of approximate 95% one-sided CBs for some commonly used methods and parametric transformed bootstrap-$t$ (PTBT) method in the case $E(r) = 3$ and $p_f = .1$.

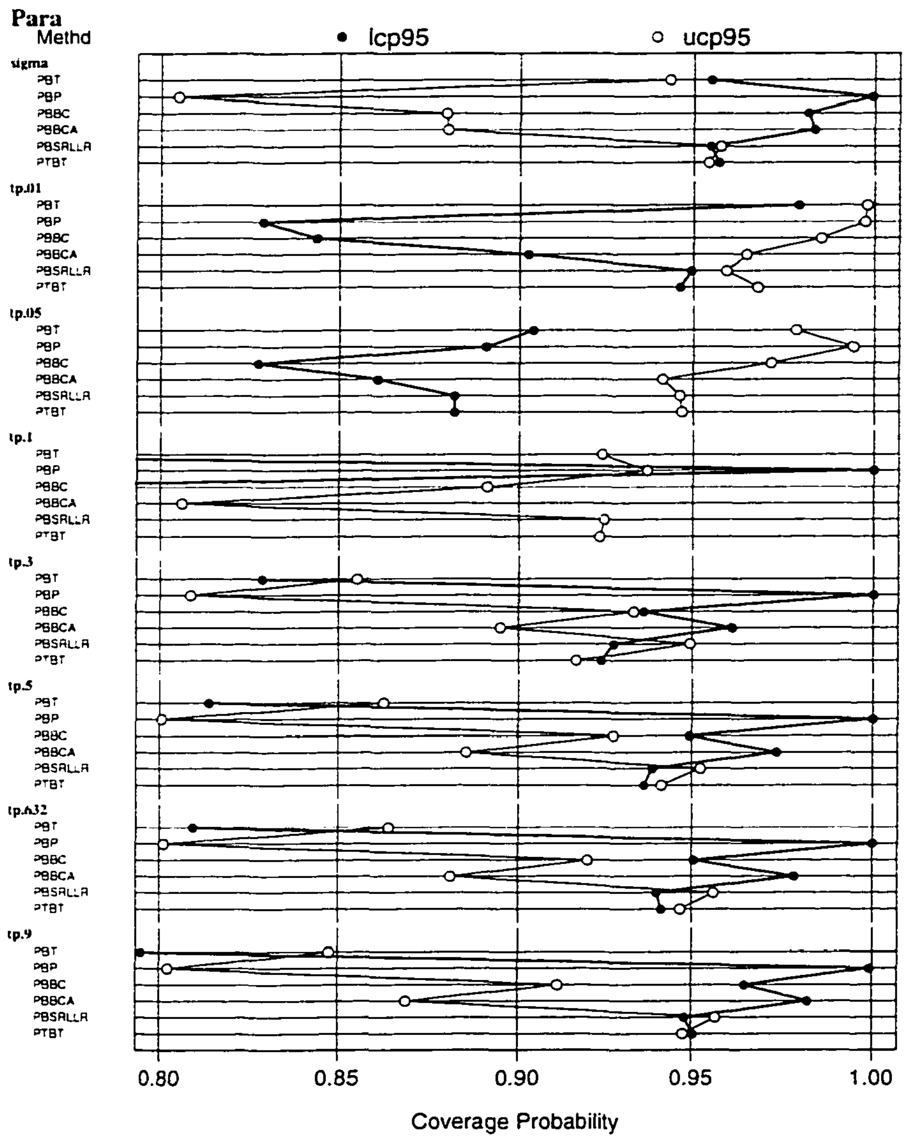



Figure 2.5 Coverage probability plot of approximate 95% one-sided CBs for bootstrap methods in the case $E(r) = 3$ and $p_f = .1$.

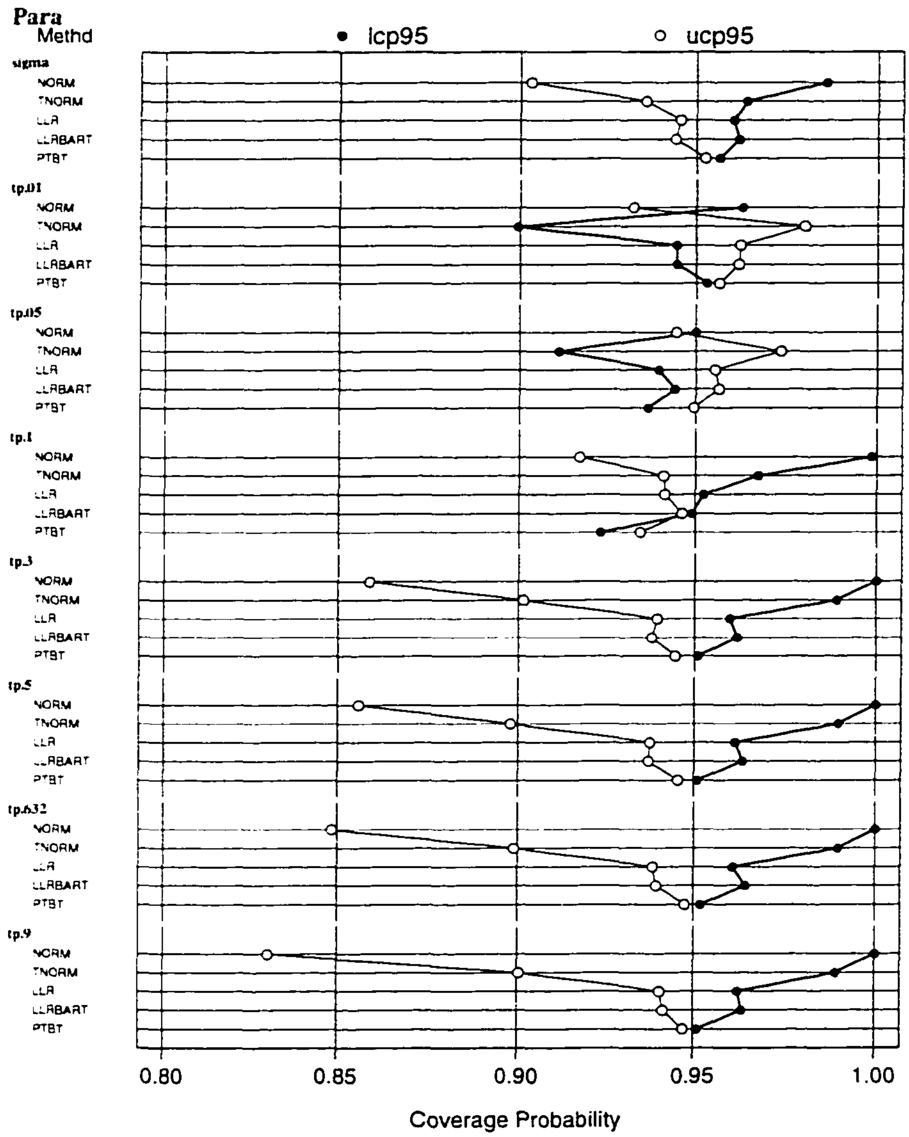



Figure 2.6 Coverage probability plot of approximate 95% one-sided CBs for some commonly used and parametric transformed bootstrap-$t$ (PTBT) method in the case $E(r) = 30$ and $p_f = .1$.

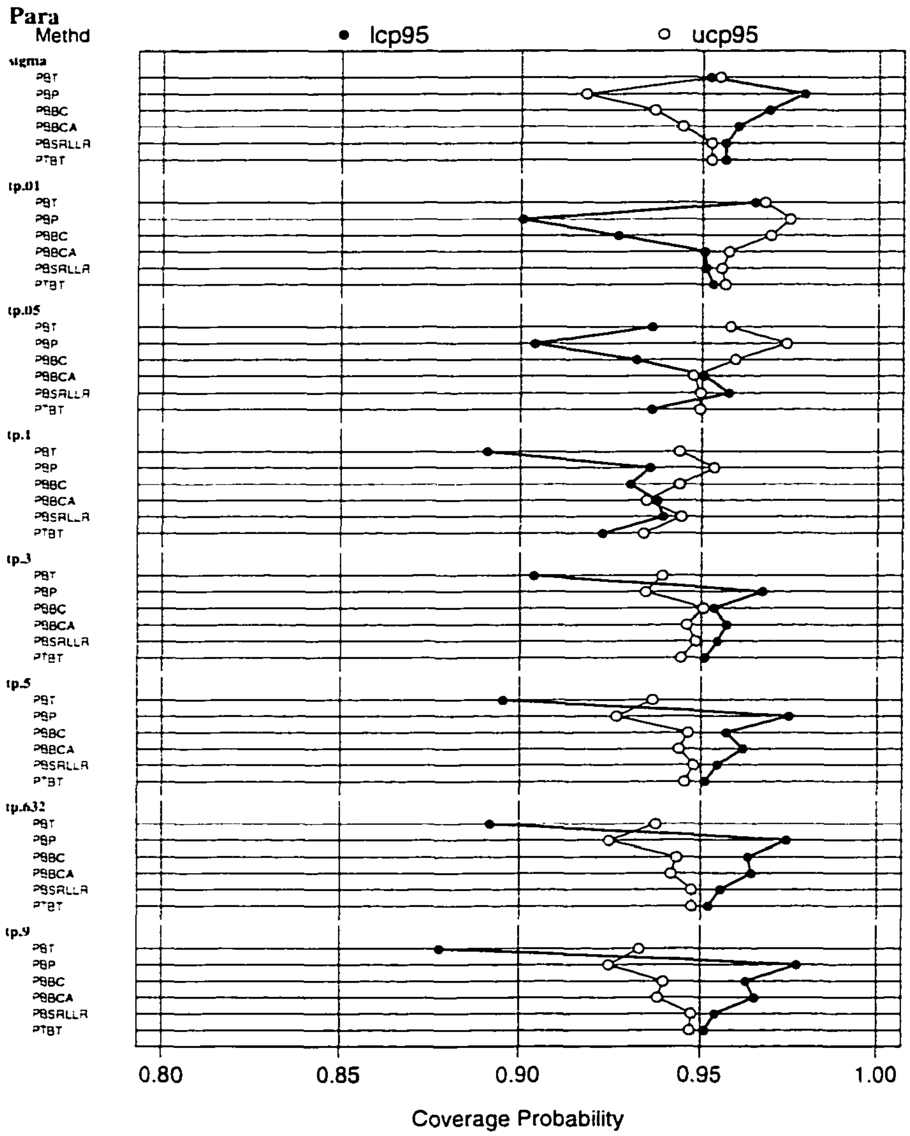



Figure 2.7 Coverage probability plot of approximate 95% one-sided CBs for bootstrap methods in the case $E(r) = 30$ and $p_f = .1$.

### 2.5.2 Two-Sided CIs

Recall that for one-sided CB's, often the CP was conservative on one side and anticonservative on the other side. With two-sided intervals, there is an averaging effect, and the overall CP approximations tend to be better. Figure 2.8 shows the CP of the two-sided 90% CIs for parameter $\sigma$ using 10 different methods for 5 cases of proportion failing. Figure 2.9 is the same type of graph as Figure 2.8 for the .1 quantile, $t_{.1}$ of the Weibull distribution.

For the parameter $\sigma$, we have following results from Figure 2.8:

- The NORM method CP is approximately equal to the nominal CP for $E(r) \geq 30$. But the TNORM method provides some improvements especially when $E(r) \geq 20$.
- The approximate CP for the LLR method is adequate for $E(r) \geq 15$. With Bartlett's correction (LLRBART), the CPs are much closer to the nominal even $E(r) = 3$.
- The approximate CP for the PBP method is adequate for $E(r) \geq 20$.
- The approximate CP for the PBBC and PBBCA method are very similar for $E(r) \geq 10$. They improve on the PBP for $E(r) \leq 20$.
- The approximate CP for the LLRBART, PBSRLLR, PBT and PTBT methods are excellent for all values of $E(r)$ and $p_f$.

For quantile parameters $t_p$, Figure 2.9 for $t_{.1}$ and similar plots for other $t_p$ values (not presented here) indicate that

- Unlike the situation for the parameter $\sigma$, the adequacy of the CP approximation depends on the expected proportion failing.
- When $p \approx p_f \leq .3$ and $E(r) \leq 20$, both NORM and TNORM are conservative. TNORM is more conservative than NORM. But for $p = p_f \geq .5$, both methods are anti-conservative. Also, when $p \neq p_f$, both methods are anti-conservative. For $p > p_f$, TNORM has CP closer to nominal than NORM. But for $p < p_f$, NORM has CP closer to nominal than TNORM.
- LLR and LLRBART have accurate CP for $E(r) \geq 15$ and $E(r) \geq 7$ respectively. For large $p_f$, LLRBART is considerably better than LLR.

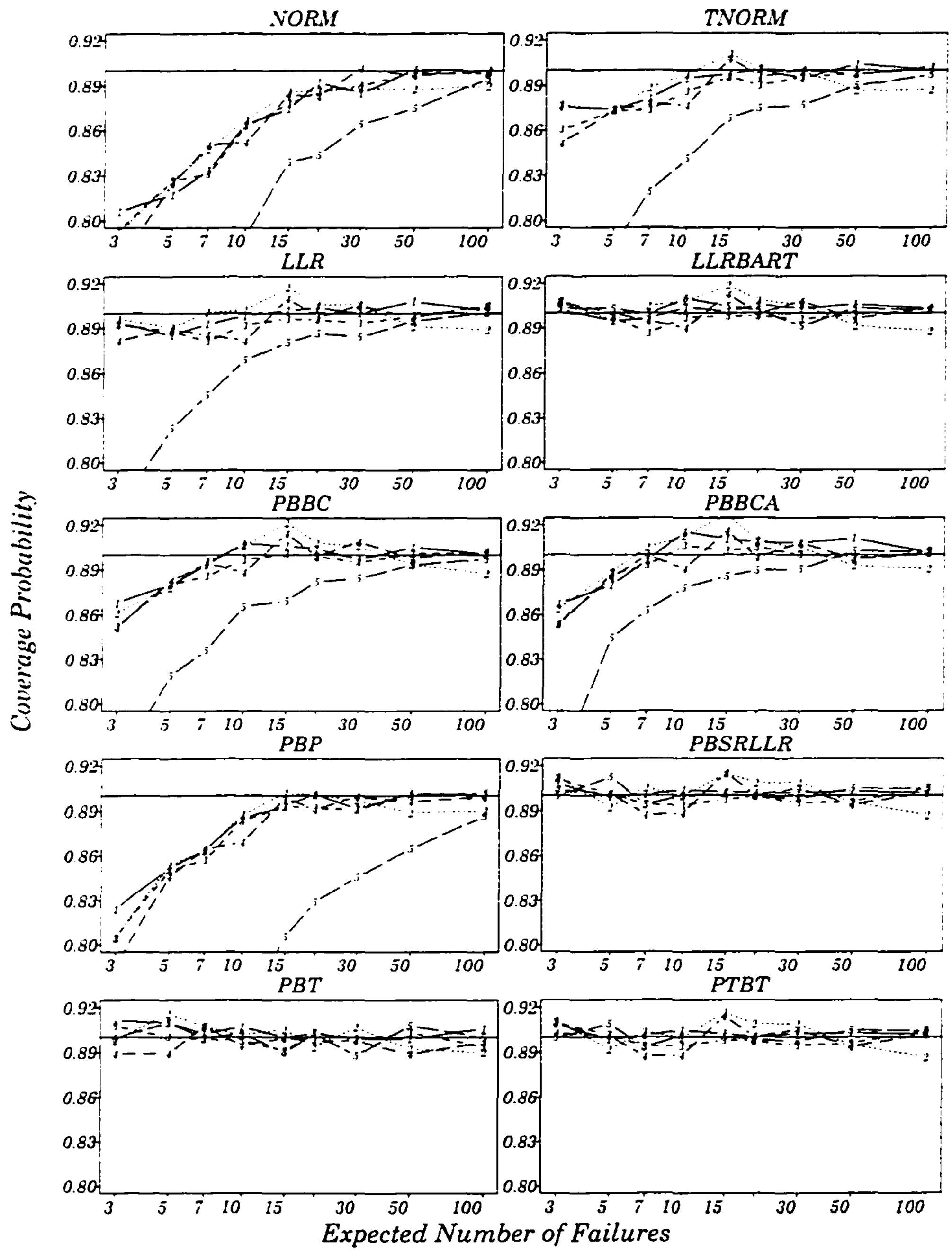



Figure 2.8 Coverage probability versus expected number of failures plot of two-sided 90% CIs for parameter $\sigma$. The numbers (1, 2, 3, 4, 5) in the lines of each plot correspond to $p_f$'s (.01, .1, .3, .5, 1).

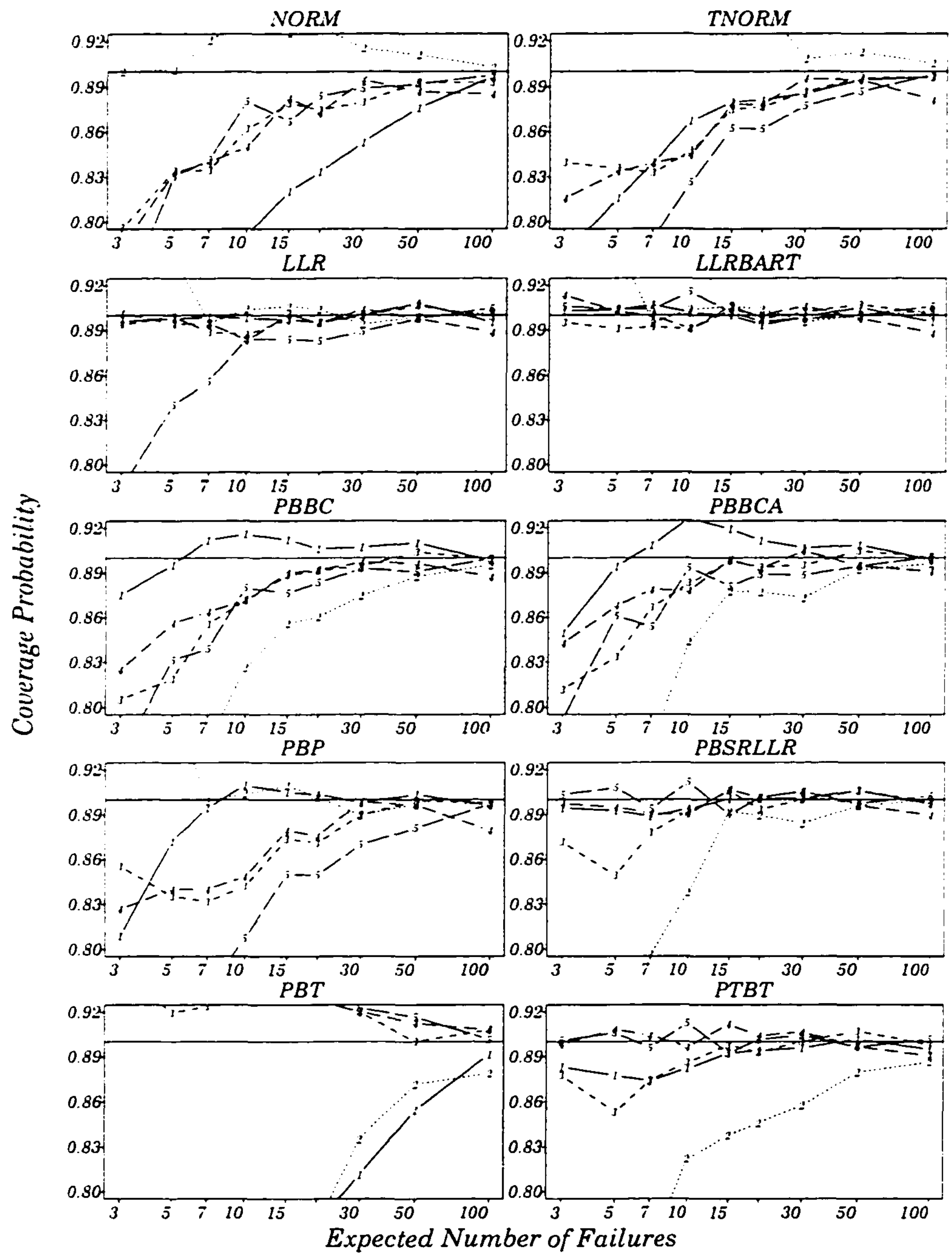



Figure 2.9 Coverage probability versus expected number of failures plot of two-sided 90% CIs for parameter $t_{.1}$. The numbers (1, 2, 3, 4, 5) in the lines of each plot correspond to $p_f$'s (.01, .1, .3, .5, 1).

- The performance of PBT is close to that of NORM and is better when $p > p_f$. But for large $E(r)$ ($> 30$), NORM is better than PBT.

- The relative performances of PBP, PBBC and PBBCA depend on both $p_f$ and $E(r)$. When $p \approx p_f$, PBBC and PBBCA tend to have lower CP than $p \neq p_f$.

- PBSRLLR and PTBT provide excellent approximations when $p > p_f$ especially when $p_f$ is small ($< .1$). When $p_f$ is close to $p$, however, both methods have CP that is lower than nominal. In this case the PBSRLLR method is better than PTBT method and provides an adequate approximation for $E(r) \geq 15$.

### 2.5.3 Expected Interval Length

Interval length is another criterion for comparing two-sided CIs. With the same coverage probability, procedures that provide shorter intervals are better. Figure 2.10 shows the average interval length of the 2000 two-sided 90% CIs for parameter $\sigma$ using 10 different methods for 5 values of $p_f$. Figure 2.11 is the same type of graph as Figure 2.10 for parameter $t_{.1}$. For parameter $\sigma$, we have the following conclusions results from Figure 2.10:

- Generally, for all different methods the CI expected length is shorter if $p_f$ is bigger. This is quite natural, as with constant $E(r)$, we have more information about the distribution for larger $p_f$.

- There is not much difference among the different procedures for $E(r) > 10$. For $E(r) \leq 10$, the order of expected lengths are :
  {NORM, PBP} < {TNORM, PBBC, PBBCA} < {LLR, LLRBART}
  < {PBSRLLR, PBT,PTBT}.

One explanation for this ordering is the anti-conservative nature of the shorter intervals.

For quantiles $t_p$, the situation is quite different from that for $\sigma$. We draw the following conclusions from Figure 2.11 and from plots for other values of $t_p$ that are not presented here:

- For the case $p \geq p_f$, the expected interval length is much wider for all different methods even when $E(r) = 20$ , due to the extrapolation in time.

- Differences in the expected length often result from differences in the CP. Intervals with more conservative CPs tend to be wider. In general, the order of the CI widths are NORM < TNORM, LLR < LLRBART, and PTBT < PBSRLLR.

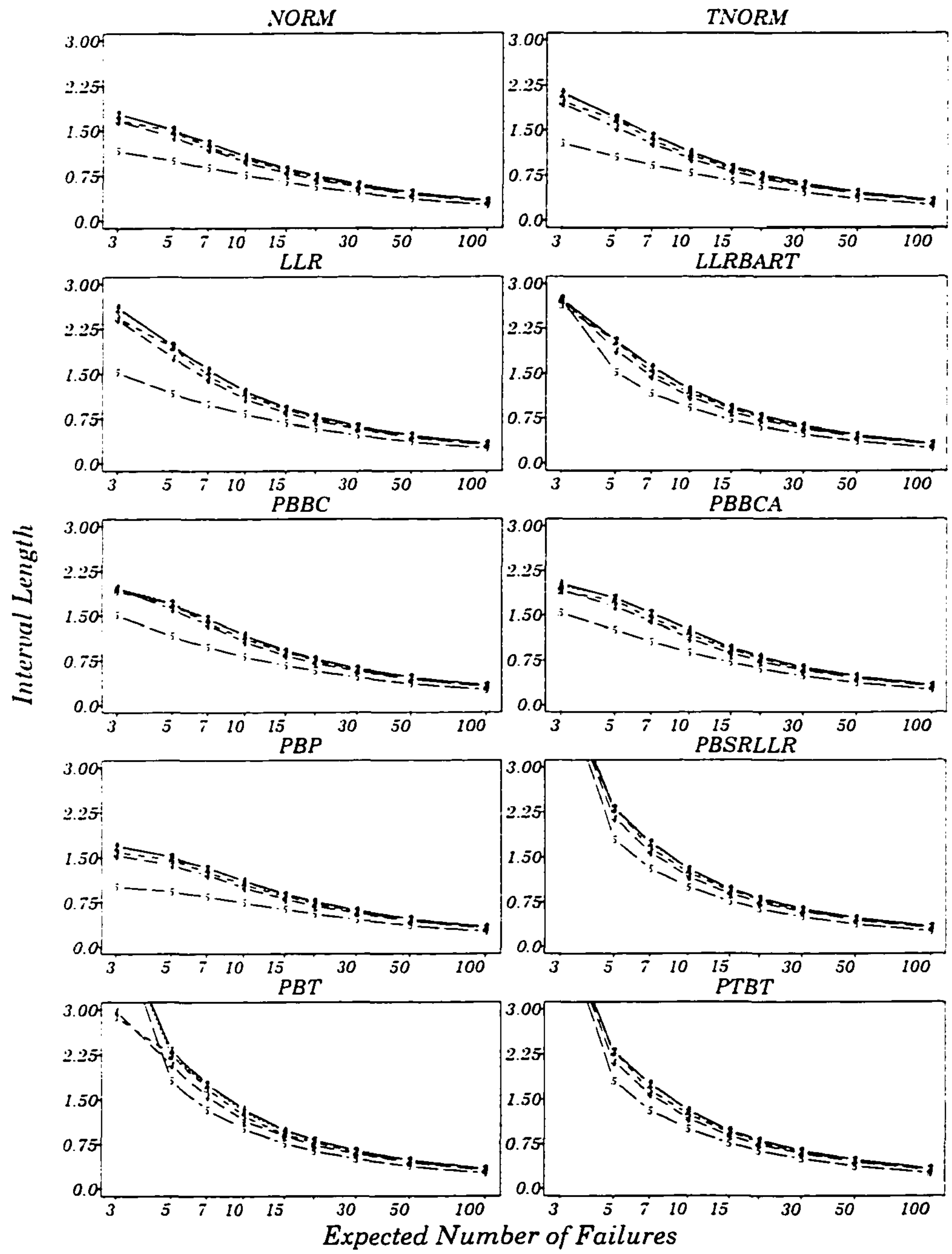



Figure 2.10 Interval length versus expected number of failures plot of two-sided 90% CIs for parameter $\sigma$. The numbers (1, 2, 3, 4, 5) in the lines of each plot correspond to $p_f$'s (.01, .1, .3, .5, 1).

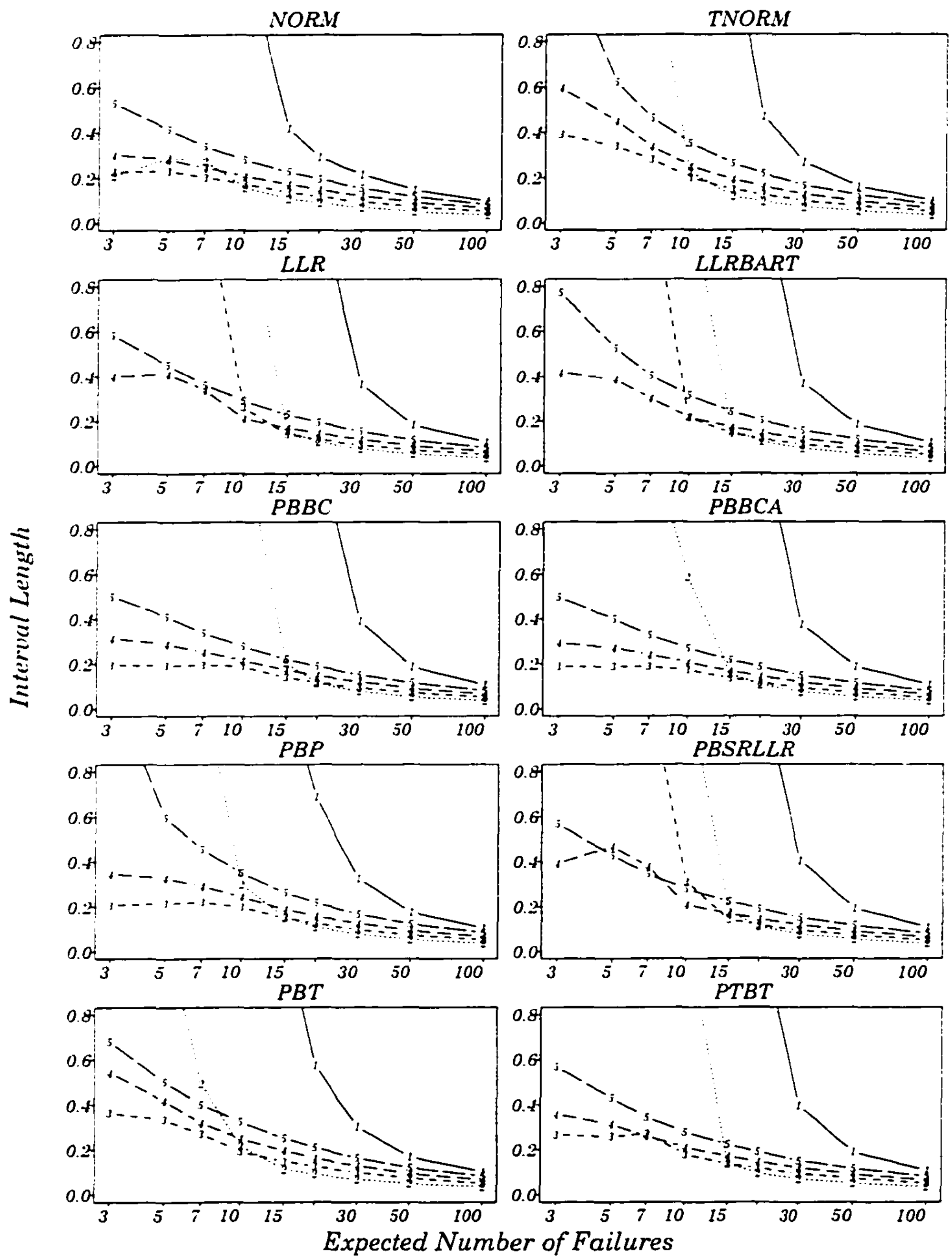



Figure 2.11 Interval length versus expected number of failures plot of two-sided 90% CIs for parameter $t_{.1}$. The numbers (1, 2, 3, 4, 5) in the lines of each plot correspond to $p_f$'s (.01, .1, .3, .5, 1).

## 2.6 Conclusions and Recommendations

Normal-approximation confidence intervals (NORM and TNORM) are still commonly used in practice and are used in many statistical software packages. Normal-approximation two-sided CIs may not be adequate when the expected number of failures is less than 50. For the one-sided case, we see that generally $E(r)$ =100 failures is still barely enough to provide a good approximation to the nominal coverage probability. If the scale parameter is of interest, the usual transformation (such as log), which makes ML estimator have range over whole real line, is suggested. Doing this usually provides a somewhat better coverage probability for any proportion failing.

Two-sided log likelihood ratio CIs have reasonably accurate coverage probability, even for expected number of failures $E(r)$ as small as 15. But the CIs are asymmetric. Individual upper and lower CIs could be somewhat conservative or anti-conservative depending on number of failures and quantiles of interest. Use of Bartlett's correction generally improves the coverage probability approximation for two-sided CIs, especially when the proportion failing is greater than .5. Those CIs are adequate even when $E(r) = 7$. But for one-sided coverage, however, the Bartlett's correction provides no improvement.

Some bootstrap methods provide better coverage probability accuracy. However, using the bootstrap-$t$ without a proper transformation may not perform any better than the normal-approximation method. It is important to use the bootstrap-$t$ procedure carefully.

The bootstrap percentile methods are easy to implement and they improve the normal-approximation method in many (but not all) cases. The accuracy of the parametric bootstrap percentile (PBP), bias corrected (PBBC) and bias-corrected accelerated (PBBCA) methods depend on the expected number of failures, the proportion failing and the parameters of interest. When the proportion failing is greater than .1, the PBBCA method has better performance than the PBBC method for quantile parameters. In heavily censored cases ($p_f < .1$), however, the PBBCA method is generally worse. This is probably due to difficulty in estimating the acceleration constant under heavy censoring.

The parametric bootstrap-$t$ with transformation (PTBT) and bootstrap signed-root log-likelihood ratio (PBSRLLR) methods provide more accurate results over all different number of failures, proportion failing and parameters of interest except for the case that parameter of interest is $t_p$ and $p$ is close to proportion failing. Moreover, the coverage probabilities are approximately symmetric, which is important when one-sided CBs are needed or when the cost of being wrong differs inportantly from one side to the other of a two-sided interval. Although the PBSRLLR method is more accurate in small samples ($E(r) < 10$), the bootstrap-$t$ with transformation requires much less computational effort than the PBSRLLR method. Inverting the signed-root log-likelihood ratio method requires repeated root finding. Also with heavy censoring, good starting values are needed to find the confidence limits and there may be numerical difficulties. However the important benefit of PBSRLLR method is that it is transformation invariant (unlike PTBT).

The CIs from both PTBT and PBSRLLR methods are wider than the normal-approximation methods (NORM and TNORM), especially when censoring is heavy. This is in part due to improving the poor accuracy of coverage probability using normal-approximations. The wider CI is as a trade off to get higher order of accuracy.

In general, when the expected number of failures is smaller than 50 (20), the likelihood ratio based methods are recommended for finding one-sided confidence bounds (two-sided confidence intervals). If one-sided CBs are required or censoring is heavy ($p_f < .1$), the PTBT and PBSRLLR methods methods are suggested except for the case when the quantity of interest is $t_p$ where $p$ is close to proportion failing. Then PBSRLLR is better than PTBT down to $E(r)=15$. When $p_f > .5$, the PBSRLLR provides accurate CP even down to $E(r) = 10$. With modern computing capabilities, the PBSRLLR method is feasible and, when appropriate software becomes available, should be considered the best practice.

## 2.7 Special Effects of Type I Censoring

In small samples, the CP from the NORM method is much more accurate if the parameter of interest is $t_p$ where $p$ is close to proportion failing $p_f$. Doganaksoy and Schmee (1993) explain that in this case $t_p$ and $\sigma$ are approximately orthogonal parameters and the NORM method benefits from this property. But unlike the NORM method, both the PTBT and PBSRLLR methods perform poorly in this situation. The possible reasons for this are $a$) the distributions of the pivotal-like statistics depend on the $p_f$ and $b$) the distributions of MLEs have a discrete or lattice component (i.e., number of failures). Figure 2.12 shows that the distributions of the pivotal-like statistics change with the value of $p_f$ and is different for quantiles $t_p$ for which $p \approx p_f$ but consistently similar for the scale parameter $\sigma$. This explains why the CP of the confidence interval for $\sigma$ is closer to nominal than the CP for confidence interval for the quantiles. Also

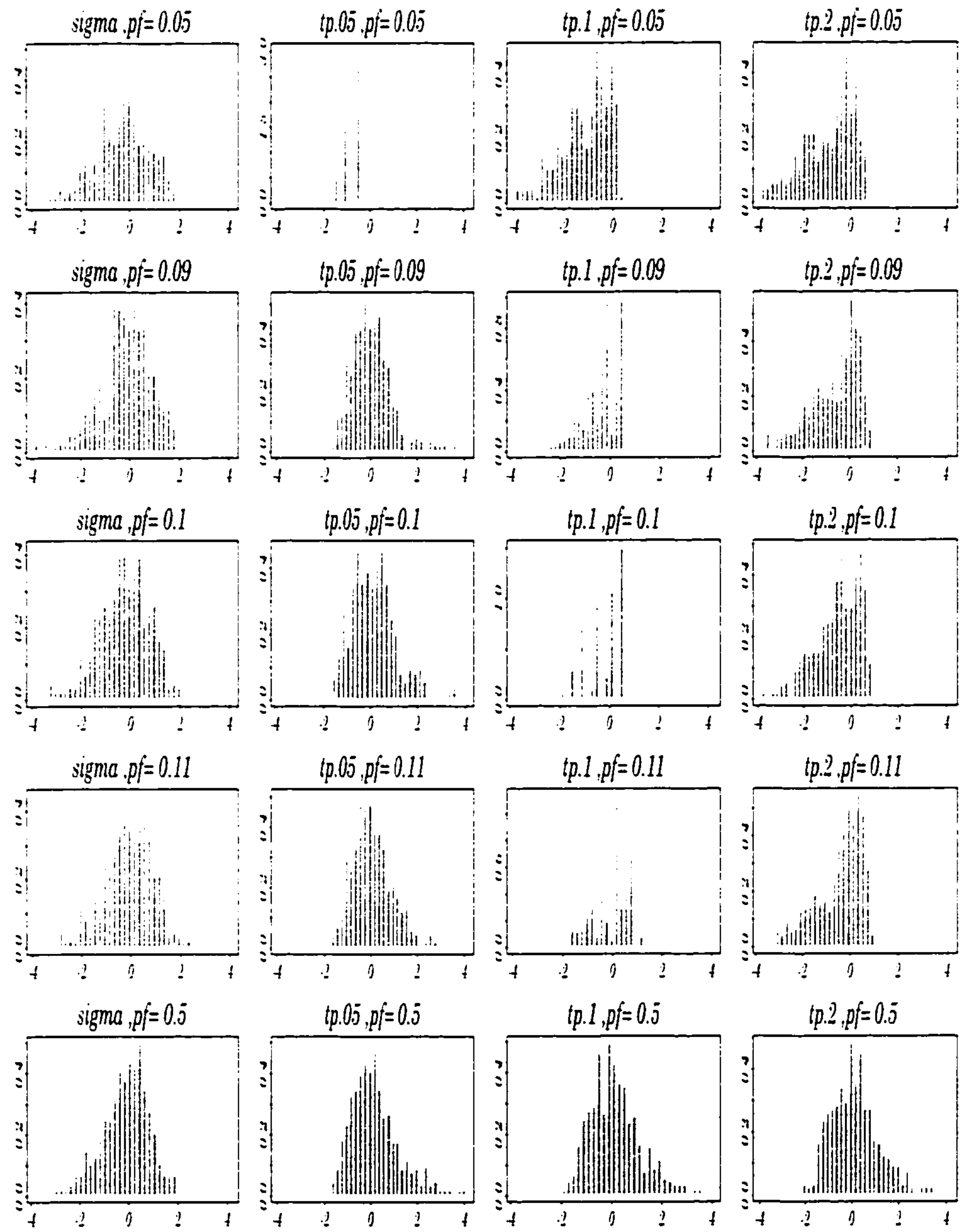



Figure 2.12 Distributions of the pivot-like statistic $(\widehat{\theta}-\theta)/\widehat{\mathrm{se}}_{\widehat{\theta}}$ in PTBT methods under Type I censoring with $p_f$ = .05, .09, .1, .11, .5. Parameters $\theta$ are $\sigma$ and $t_p$, for $p$ = .05, .1, and .2.

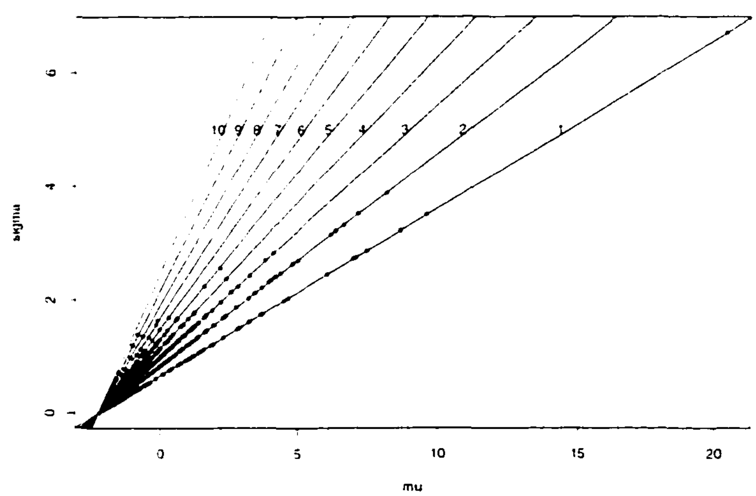



Figure 2.13 Plotted points are values of $(\widehat{\mu}, \widehat{\sigma})$ from 500 MLEs from Weibull distribution data with $p_f = .1$ and sample size 30 (so $E(r) = 3$). The lines are $\log(t_{.1}) = \mu + \Phi^{-1}(r/30)\sigma$, $r = 1, 2, \ldots, 10$. $\Phi$ the Weibull cdf, where $\mu = 0$, $\sigma = 1$.

there is a strong discrete-like behavior in the sampling distribution of some statistics (e.g., $\widehat{t_p}$) when $p$ of $t_p$ is close to $p_f$. With Type I censored data, if $r/n$ is small (<.5) then $\log(t_c) \approx \widehat{\mu} + \Phi^{-1}(r/n)\widehat{\sigma}$, where $t_c$ the censor time, $r$ the number of failures and $n$ the sample size. This is illustrated in Figure 2.13.

Robinson (1983) applied a parametric bootstrap method (see the Appendix A.2) to find CIs for multiple time-censored progressive data. This method (like PTBT) is exact when data are complete or Type II censored. Since multiple time-censored progressive data contain several censoring times, there is no discrete-like behavior in the MLEs like that in Type I censoring. For this reason the CP with multiple time censoring is close to the nominal over all of the different cases. For Type I censored case, however, the coverage probability of Robinson's method is less accurate than that of the transformed bootstrap-$t$ (PTBT) method.

## 2.8 Discussion and Directions for Future Research

It is most common that life tests result in Type I censored data. Because there are no known exact confidence interval methods for Type I censored data, this paper provides a detailed comparison of methods for constructing approximate confidence intervals. These methods range from the most commonly used large-sample normal-approximation methods to the more modern computationally-intensive likelihood and simulation-based methods. Because opposite lower and upper bounds of a two-sided confidence interval tend to have conservative versus anti-conservative coverage probabilities, the effect of averaging often results in reasonably adequate coverage-probability approximations for two-sided confidence intervals in situations with moderately large sample sizes. Our results show, however, that for moderate amounts of censoring and one-sided bounds (most commonly used in practical applications in the physical and engineering sciences as well as other areas of application) the simple normal-approximation (NORM and TNORM) methods provide only crude approximations even when the expected number of failures is as large as 100.

Appropriate computationally-intensive methods provide important improvements. In particular, likelihood-based methods, even when calibrated with the large sample chi-square distribution approximation (e.g., the LR method), generally provide important improvements. Calibrating the LR CIs by simulation (see the Appendix A.2) does not address the asymmetry problem and results in inaccurate one-sided bounds. Calibrating the individual tails of a likelihood-based interval with simulation (i.e., the PBSRLLR method) provides important improvements in coverage probability accuracy. even for small E($r$). for all but one exceptional situation (i.e.. inferences at times near to the censoring time or quantiles near the proportion censoring with $E(r) \leq 10$). The transformed bootstrap-$t$ procedure provides a computationally simpler method. but one needs to be careful in the specification of the transformation to be used.

In addition to providing guidance for practical applications. our results suggest the following avenues for further research.

1. Our study leaves unanswered the question of what one should do when making inferences in the exceptional case when the failure number is down to 10. We see no easy solution to this problem. Some possibilities include

   - Extending the censoring time of the life test to be safely and sufficiently beyond the time point (or proportion failing) of interest. This requires prior knowledge of the failure-time distribution which is not generally available.
   - Design life test experiments to result in Type II censored data. In this case. exact confidence interval procedures are available. but experimenters generally have to deal with time constraints in life testing and thus there may be resistance to such life test plans. On the other hand. Type II censoring provides important control over the amount of information that a life-test experiment will provide.
   - Design life test experiments to result in multiple time-censoring (where the results of Robinson (1983) suggest that excellent large sample approximations are available from computationally intensive methods). In this case. constraints on time or number of units available for testing may also lead to resistance to such life test plans.
   - If none of the above is possible (e.g.. for reasons given above or because the experiment has already been completed) it might be possible to make use of nonparametric methods (where conservative confidence intervals or bounds may be available if there is a sufficient amount of data).

2. Our study has focused on the Weibull distribution. It would be of interest to replicate the study for other distributions. We would expect very similar results for other log-locations-scale distributions such as the lognormal and the loglogistic distributions.

3. It would be of interest to extend this study to other censored-data situations that arise in applications, including regression analysis and the analysis of accelerated life test data, more complicated censoring schemes like interval censoring and random censoring, simultaneous confidence interval and bounds, intervals to compare two different grouped, and so on.

4. The LLRBART is second-order accurate for two-sided CI using Type I censored data (Jensen 1993). Both PTBT and PBSRLLR methods are better than LLBART in one-sided cases. Simulation results also suggest that PBSRLLR is better than PTBT with smaller sample sizes. This suggests that higher-order asymptotics would show a difference between these different methods. This could be explored.

## References


Barndorff-Nielsen, O. E. (1986). "Inference on full or partial parameters based on the standardized signed log likelihood ratio." *Biometrika*, **73**, 307-322.

Barndorff-Nielsen, O. E. (1991). "Modified signed log likelihood ratio." *Biometrika*, **78**, 557-563.

Barndorff-Nielsen, O. E., and Cox, D. R. (1994). *Inference and Asymptotics*. London: Chapman & Hall

Bartlett, M. S. (1937). "Properties of sufficiency and statistical tests." *Proceedings of Royal Society of London* A **160**, 268-282.

Casella, G. and Berger, R. L. (1990). *Statistical Inference*. Belmont: Duxbury Press.

Doganaksoy, N. (1995). "Likelihood ratio confidence intervals in life-data analysis." *Recent Advances in Life-Testing and Reliability*, N. Balakrishnan (Editor). Boca Raton: CRC Press. Chapter 20.

Doganaksoy, N., and Schmee, J. (1993). "Orthogonal parameters with censored data." *Communications in Statistics, Part A, Theory and Methods*, **22**, 669-685.

Efron, B. (1981). "Nonparametric standard errors and confidence intervals (with discussion)." *The Canadian Journal of Statistics*, **9**, 139-172.

Efron, B. (1982). *The Jackknife, the Bootstrap and Other Resampling Plans*, Society for Industrial and Applied Mathematical, Philadelphia.

Efron, B. (1987). "Better bootstrap confidence intervals (with discussion)." *Journal of the American Statistical Association*, **82**, 171-200.

Efron, B., and Tibshirani, R. J. (1993). *An Introduction to the Bootstrap*. New York: Chapman & Hall.

Hall, P. (1987). "On the bootstrap and likelihood-based confidence regions." *Biometrika*, **74**, 481-493.

Hall, P. (1992). *The Bootstrap and Edgeworth Expansion*. New York: Springer-Verlag.

Jensen, J. L. (1993). "A historical sketch and some new results on the improved log likelihood ratio statistic." *Scandinavian Journal of Statistics, Theory and Applications*, **20**, 1-15.

Lawless, J. F. (1982). *Statistical Models and Methods for Lifetime Data*. New York: John Wiley & Sons.

Lawless, J. F. (1983). "Statistical methods in reliability." *Technometrics*, **25**, 305-316.

Meeker, W. Q., and Escobar, L. A. (1994). "Maximum likelihood methods for fitting parametric statistical models to censored and truncated data." *Probabilistic and Statistical Methods in the Physical Sciences*, edited by J. Stanford and S. Vardeman. New York: Academic Press.

Mood, A. M., Graybill, F. A., and Bose, D. C. (1974). *Introduction of the Theory of Statistics*, 3rd ed., New York: McGraw-Hill.

Nelson, W. (1982). *Applied Life Data Analysis*. New York: John Wiley & Sons.

Ostrouchov, G., and Meeker, W. Q., Jr. (1988). "Accuracy of approximate confidence bounds computed from interval censored Weibull and lognormal data." *Journal of Statistical Computation and Simulation*, **29**, 43-76.

Peddada, S. D., and Patwardhan, G. (1992). "Qualms about $BC_a$ bootstrap confidence intervals." *Statistics & Probability Letters*, **15**, 77-83.

Piegorsch, W. W. (1987). "Performance of likelihood-based interval estimates for two-parameter exponential samples subject to Type I censoring." *Technometrics*, **29**, 41-49.

Robinson, J. A. (1983). "Bootstrap confidence intervals in location-scale models with progressive censoring." *Technometrics*, **25**, 179-187.

Schenker, N., and Patwardhan, G. (1985). "Qualms about bootstrap confidence intervals." *Journal of the American Statistical Association*, **80**, 360-361.

Serfling, R. J. (1980) *Approximation Theorems of Mathematical Statistics*. New York: John Wiley & Sons.

Shao, J., and Tu, D. (1995). *The Jackknife and Bootstrap*. New York: Springer-Verlag.

Vander Wiel, S. A., and Meeker, W. Q. (1990). "Accuracy of approximate confidence bounds using censored Weibull regression data from accelerated life tests." *IEEE Transactions on Reliability*, **39**, 346-351.

# 3 BOOTSTRAP LIKELIHOOD RATIO STATISTICS

A paper to be submitted to Scandinavian Journal of Statistics

Shuen-Lin Jeng and William Q. Meeker

## Abstract

Much research has been done to study the asymptotic distributions of likelihood ratio statistics. Most of this research has focused on the situation in which the underling distribution is continuous (especially, parametric families) or discrete (e.g., the empirical distribution). In this paper we consider the situation in which the data are censored and the distribution of the likelihood ratio statistic is a mixture of continuous and discrete distributions. Jensen (1993) shows that under this situation the distribution of Bartlett-adjusted likelihood ratio statistics can be approximated by a $\chi^2$ distribution up to second order accuracy $[O(1/n)]$. This result can be used to provide a second order accurate procedure for constructing confidence intervals. However, if the one-sided confidence bound is of interest, the coverage probability of a procedure is usually only first order accurate when using the $\chi^2$ approximation. We extend the results from Jensen (1993) and show that the distribution of signed square root likelihood ratio statistics can be approximated by its bootstrap distribution up to second order accuracy. Similar results apply to likelihood ratio statistics with or without a Bartlett correction. We use a simulation study to investigate the adequacy of the approximation provided by the theoretical result. We compare the finite-sample coverage probability of several competing confidence interval procedures based on the two parameters Weibull model. The bootstrap-t and $BC_a$ methods are second order accurate when the data are complete. Our simulation results show that the methods based on bootstrap signed square root likelihood ratio statistics and its modification outperform the bootstrap $t$ and $BC_a$ methods in constructing one-sided confidence bounds when the data are Type I censored.

**Keywords:** Bartlett correction, confidence interval, life data, likelihood ratio, one-sided confidence bound, signed square root likelihood ratio, parametric bootstrap, Type I censoring.

## 3.1 Introduction

### 3.1.1 Motivation

The asymptotic distributions of likelihood ratio statistics had been studied for decades. Most previous work has focused on the situations in which the underling distribution is continuous (especially parametric families) or discrete (e.g. empirical distributions). The log likelihood ratio statistic usually provides more accurate approximate inferences than the more commonly used studentized maximum likelihood estimators (see for example, Doganaksoy and Schmee (1993) and Jeng and Meeker (1998)). For finding one-sided confidence bounds, the signed square root of log likelihood ratio (SRLLR) statistic is commonly used. However, even for complete continuous data, generally the SRLLR statistics is approximated by the standard normal distribution only to first order $[O(1/\sqrt{n})]$ (Barndorff-Nelsen and Cox (1994)). With time censored data, the distribution of a likelihood ratio statistic is a mixture of continuous and discrete distributions. Jensen (1993) derives an Edgeworth expansion of log likelihood ratio (LLR) statistic when its underlying distribution is partly discrete. The bootstrap is a general resampling or simulation procedure to find an approximate sampling distribution. In this paper we extend the results from Jensen (1993) and we show that, under some regularity conditions that apply to complete and censored data, the distribution of the SRLLR statistic and the LLR statistics, with or without a Bartlett correction, can be approximated up to second order accuracy $[O(1/n)]$ by using the bootstrap procedure.

### 3.1.2 Literature Review

Let $l(x;\theta), \theta = (\theta^{(1)}, \theta^{(2)}) = (\theta_1, \ldots, \theta_{k_1}, \theta_{k_1+1}, \ldots, \theta_k)$, be the log likelihood function for a single observation $x$ where $\theta^{(2)}$ is the parameter of primary interest and $\theta^{(1)}$ is a vector of nuisance parameter. When there are $n$ observations, define $\bar{l}_n$ as

$$\bar{l}_n(\theta) = \frac{1}{n}\sum_{1}^{n} l(x_i;\theta). \tag{3.1}$$

where $x_i$ is the data for the observation $i$. Let $\widehat{\theta} = (\widehat{\theta}_1, \ldots, \widehat{\theta}_k)$ and $\tilde{\theta} = (\tilde{\theta}_1, \ldots, \tilde{\theta}_{k_1}, \theta_{(k_1+1)0}, \ldots, \theta_{k0}) = (\tilde{\theta}^{(1)}, \theta_0^{(2)})$ be the maximum likelihood estimates for the full model parameter vector $\theta$ and for the restricted model parameter $\theta^{(2)} = \theta_0^{(2)}$, respectively. Then the log likelihood ratio statistic is

$$W_n = 2n[\bar{l}_n(\widehat{\theta}) - \bar{l}_n(\check{\theta})], \tag{3.2}$$

and under standard regularity conditions (e.g., Lehmann 1986) $W_n$ is asymptotically $\chi^2_{(k-k_1)}$, where $\chi^2_f$ denotes a chi-square distribution with degree of freedom $f$. The signed square root log likelihood ratio (SRLLR) statistic for testing a scalar parameter (or a scalar function of the parameter so that $k_1 = k - 1$) $\theta_k = \theta_{k0}$ is

$$R_n = \text{sign}(\widehat{\theta}_k - \theta_{k0})\sqrt{W_n}, \tag{3.3}$$

and the distribution of $R_n$ is asymptotically standard normal.

The distribution of the log likelihood statistic for i.i.d complete data has been described in a number of publications. Box (1949) derives an infinite series for the distribution of $W_n$ in terms of the $\chi^2$ distribution and with terms decreasing in powers of $1/n$. Lawley (1956) derives the Bartlett correction term for $W_n$. Hayakawa (1977) gives an asymptotic expansion of the distribution of $W_n$ with the i.i.d complete data as the following

$$\Pr(W_n \le w) = \Pr(\chi^2_{k-k_1} \le w) + \frac{1}{24n}\Big\{A_2 \Pr(\chi^2_{k-k_1+4} \le w) - (2A_2 - A_1)\Pr(\chi^2_{k-k_1+2} \le w)$$
$$+ (A_2 - A_1)\Pr(\chi^2_{k-k_1} \le w)\Big\} + o\left(\frac{1}{n}\right). \tag{3.4}$$

where $\chi^2_k$ is a chi-square random variable with $k$ degrees of freedom, $A_1$ and $A_2$ are functions of $\theta$, and $A_2 = 0$ when testing $\theta = \theta_0$. Chandra and Ghosh (1979) give an asymptotic expansion of $W_n$ to order $o(1/n)$ in terms of the maximum likelihood estimator of $\theta$ as well as the second, third and fourth derivatives of $n\bar{l}_n(\theta)$.

Lawley (1956) gives a series expansion of $R_n$ in terms of the first four derivatives of $n\bar{l}_n(\theta)$ and their expectations. When $\theta$ is a scalar, McCullagh (1984) argues that by conditioning on a second-order locally ancillary statistic, the procedure based on $R_n$ plus the standardized skewness of $n\partial\bar{l}_n/\partial\theta$ has the desired coverage probability to the order $O(1/n)$. Therefore the procedure, unconditionally, has error $O(1/n)$ in each tail. Efron (1985) shows that $R_n$ is asymptotically normal to order $O(1/n^{3/2})$. Barndorff-Nelson (1986, 1991) verifies that a particular modification of $R_n$ follows standard normal distribution to order $O(1/n^{3/2})$ conditionally on an appropriate ancillary, and hence also unconditionally. Nishii and Yanagimoto (1993) provide an asymptotic expansion of studendized $R_n$ up to third order for distributions in the exponential family.

For censored data, the usual arguments for finding a formal Edgeworth expansion are no longer valid. The order of accuracy in the results mentioned above could be different. Jensen (1987, 1989) establishes the Edgeworth expansion for a smooth function of the mean of some statistics when the underling distribution is partly discrete. He first derives the expansion conditional on the discrete part and then integrates over the discrete component to obtain the unconditional Edgeworth expansion. Babu (1991) calculates the Edgeworth expansion for statistics that are functions of lattice and non-lattice variables for the case that the lattice variable is only one dimensional. Jensen (1993) shows that $W_n$ has a three-term Edgeworth expansion.

A large number of bootstrap methods have been suggested for testing or finding confidence intervals (Hall 1992, Efron 1993, Shao and Tu 1995). The theoretical arguments for the accuracy of these methods are mostly derived under the assumption of complete data. For Type I censored data, some bootstrap methods can be much less accurate, especially for one-sided confidence intervals and small expected number failing (see Jeng and Meeker, 1998). Datta (1992) establishes a continuous version of classical Edgeworth expansions for both non-lattice and lattice distributions and uses this to unify both non-parametric and parametric bootstrap methods of a studenized statistic up to order $O(1/\sqrt{n})$. Datta (1992) gives an example that bootstrap-$t$ method is first order accurate $[O(1/\sqrt{n})]$ for the Type I censored data with the exponential distribution.

In Section 3.2, we establish a result that the distribution of the SRLLR and LLR statistics with or without a Bartlett correction can be approximated to order $O(1/n)$ by its bootstrap distribution when the underling distribution is partly discrete. Section 3.3 gives examples for using the theorems in Section 3.2. A simulation study is used to compare the finite sample properties of several different methods. Section 3.4 concludes with a summary and discussion of some possible areas for future research. In order to keep this paper self-contained, Appendix B contains statements of some important results from the literature.

## 3.2 Theorems and Results

We want to establish that the distributions of the SRLLR and LLR statistics can be approximated by the distributions of their bootstrap version to the order $O(1/n)$ for complete and censored data. We use an approach that has two stages.

1. Express the likelihood ratio statistic as a function of $\sqrt{n}g(S_n/n)$ plus some higher order error terms, where $S_n = (X_n, Y_n)$, $X_n$ is a continuous variable with mean zero in $\mathbb{R}^{r_1}$, $Y_n$ is a lattice variable with mean $\mu_n$ in $\mathbb{R}^{r_2}$ having minimal lattice $\mathbb{Z}^{r_2}$, and $g$ is a smooth function. Then find an Edgeworth expansion for the distribution of the statistic $\sqrt{n}g(S_n/n)$. This will establish that the likelihood ratio statistics has the same Edgeworth expansion up to a certain order.

2. Prove that the Edgeworth expansion of the likelihood ratio statistic can be approximated by its bootstrap version up to second order accuracy.

The work in the first stage is essentially done by Jensen (1987, 1989, 1993). For completeness of this paper we state some of his results (mostly in Section 3.2.1 and Appendix B). Then, based on Jensen's theorem we establish our main result in Section 3.2.2.

### 3.2.1 Likelihood Ratio Statistics

For a function of $\theta \in \mathbb{R}^k$ we denote by $\partial^\nu$ the partial derivative $\partial^{|\nu|}/(\partial\theta_1^{\nu_1}\ldots\partial\theta_k^{\nu_k})$ where $\nu \in \mathbb{N}^k$, $|\nu| = \sum \nu_i$ and $\nu! = \nu_1!\ldots\nu_k!$. When $|\nu| = 1$ we write $\partial_i$ instead of $\partial^\nu$ to denote a partial derivative w.r.t. $\theta_i$. Let $X_1, X_2, \ldots$ be an i.i.d. sequence of real value random variables with common distribution $P_\theta$, where $\theta$ belongs to an open subset $\Theta$ of $\mathbb{R}^p$. Let $l(x;\theta)$ is $(\mathfrak{X}, \mathfrak{A})$ measurable function w.r.t. some measure $\mu$. Denote the cdf and density of $P_\theta$ by $F(x;\theta)$ and $f(x;\theta)$, respectively. Typically $l$ can be the logarithm of the likelihood function of an observation. For example when the data are not censored, $l(X_i;\theta) = \log[f(X_i;\theta)]$. With single Type I censoring at censor time $t_c$, we have

$$l(X_i;\theta) = \log\{f(X_i;\theta)^{\delta_i}[1 - F(X_i;\theta)]^{1-\delta_i}\}. \tag{3.5}$$

where $\delta_i = 1$, if $X_i \le t_c$ (a failure) and $\delta_i = 0$, if $X_i > t_c$ (a censored observation), $i = 1, \ldots$.

Let the true parameter value be $\theta = \theta_0$ and $\nu$ be a $k$ dimensional nonnegative integer vector. We shall use the following regularity conditions with $s(\ge 3)$ a fixed integer.

Let the mean of the log likelihood functions $\bar{l}_n$ be defined as in (3.1). Then the log likelihood ratio (LLR) statistic $W_n$ is defined as in Equation (3.2). The signed square root of the log likelihood ratio (SRLLR) statistic is defined as in Equation (3.3).

The following are the regularity conditions for the likelihood function $l$.

(A) Conditions:

(A1) For each $\nu$, $1 \leq |\nu| \leq s+1$, $l(x:\theta)$ has a $\nu$-th partial derivative $\partial^\nu l(x:\theta)$ with respect to $\theta$ on $\mathfrak{X} \times \mathfrak{A}$.

(A2) For each $\nu$, $1 \leq |\nu| \leq s$, $\mathrm{E}\left[|\partial^\nu l(X_1:\theta_0)|\right] < \infty$ and there exists $a_1 > 0$ such that for each $\nu$, $|\nu| = s+1$,

$$\mathrm{E}\left[\sup_{|\theta-\theta_0|<a_1} \{|\partial^\nu l(X_1:\theta)|\}^s\right] < \infty.$$

(A3) $\mathrm{E}\left[\partial_i l(X_1:\theta_0)\right] = 0$ for $i = 1, \ldots, k$, and the $k \times k$ matrices

$$I(\theta_0) = \{-\mathrm{E}[\partial_i\partial_j l(X_1:\theta_0)]\} \tag{3.6}$$

$$D(\theta_0) = \{\mathrm{E}[\partial_i l(X_1:\theta_0)\partial_j l(X_1:\theta_0)]\}$$

are non-singular, and $I(\theta_0) = D(\theta_0)$.

Define $Z_i^{[\nu]} = \partial^\nu l(X_i:\theta_0)$ and let $Z_i = (Z_i^{[\nu]})_{1\leq|\nu|\leq s}$ be the vector with coordinates indexed by the $\nu$'s. The dimension of $Z_i$ is $m = \sum_{r=1}^{s} \binom{k+r-1}{r}$, and we arrange $Z_i$ values such that the first $k$ coordinates of $Z_i$ are those with indices $\nu = \epsilon_j \in \mathbb{N}^k$ that have their $j$-th coordinate equal to one and the rest equal to zero. Some of the coordinates of $Z_i$ may be linearly dependent and we write

$$Z_i = \tilde{Z}_i A, \tag{3.7}$$

where $\tilde{Z}_i$ has dimension $m_0 \leq m$. Then $\tilde{Z}_i$ has linearly independent coordinates of which the first $m_1$ are continuous variables and the remaining $m_2 = m_0 - m_1$ are lattice variables with minimal lattice $\mathbb{Z}^{m_2}$. We will write $\tilde{Z}_i = (\tilde{Z}_i^{(1)}, \tilde{Z}_i^{(2)})$, where $\tilde{Z}_i^{(1)}$ are the first $m_1$ coordinates and $\tilde{Z}_i^{(2)}$ are the last $m_2$ coordinates.

(B) Conditions:

(B1) $\mathrm{E}\left[|\tilde{Z}_1|^{\max\{2s+1, m_1+1\}}\right] < \infty$, $\mathrm{E}\left[|\tilde{Z}_1^{(2)}|^{\max\{2s+1, m_1+1, m_2+1\}}\right] < \infty$ and for all $\varepsilon > 0$ there exists a $\rho < 1$ such that

$$\left|\mathrm{E}\left[\exp(it\cdot\tilde{Z}_1^{(1)} + iv\cdot\tilde{Z}_1^{(2)})\right]\right| \leq \rho$$

for $|v_j| \leq \pi$, $j = 1, \ldots, m_2$ and one of $|t| > \varepsilon$ or $|v| > \varepsilon$ being fulfilled.

(B2) The $m_1 \times k$ matrix $A^{(11)}$ has full rank, where $A^{(11)}$ is the upper left hand corner of $A$.

(B3) The $m_1 \times (k-k_1)$ matrix $(A^{(1)}I(\theta_0)^{-1/2})^{(12)}$ has full rank, where $A^{(1)}$ is the matrix consisting of the first $k$ columns of $A$, the lower triangular matrix $I(\theta_0)^{-1/2}$ is the Cholesky factorization of $I(\theta_0)^{-1}$, and $(A^{(1)}I(\theta_0)^{-1/2})^{(12)}$ is the $m_1 \times (k-k_1)$ matrix of the first $m_1$ rows and columns $(k_1+1,\ldots,k)$ of $A^{(1)}I(\theta_0)$.

Condition $(B1)$ is called a uniform Cramer condition, and is required to establish an Edgeworth expansion for the continuous part given the lattice part $\tilde{Z}_i^{(2)}$. Condition $(B2)$ is used to assure that the part corresponding to the parameter $\theta^{(2)}$ in a first order Taylor approximation of the statistic, $\sqrt{n}g(S_n/n)$, depends on the continuous part $\tilde{Z}_i^{(1)}$. Condition $(B3)$ is used to assure the invariant property of the reparameterization.

Jensen (1993) gives a proposition that can be used to check the Condition (B1). We state it as the following.

**Proposition 1 (Jensen (1993), Proposition 2.3)** . *Let $(X,Y) \in \mathbb{R}^{m_1} \times \{0,1\}^{m_2}$ and assume that the random vector has a continuous component with density $f(x,y)$ with respect to product measure of Lebesgue measure and counting measure. Assume that there exist $c>0$ and $a>0$ such that*

$$f(x,y) > c, \quad \text{for } (x,y) \in (-a,a)^{m_1} \times \{0,1\}^{m_2}.$$

*Then for all $\epsilon > 0$ there exists a $\rho < 1$ such that*

$$|\mathrm{E}[\exp(it \cdot X + iv \cdot Y)]| < \rho$$

*for $|v| \le \pi$, $j = 1,\ldots,m_2$ and one of $|t| > \epsilon$ or $|v| > \epsilon$ being fulfilled.*

**Proof.** See Jensen (1993), Proposition 2.3.

The following lemma given by Jensen (1993) provides an asymptotic expansion of the SRLLR statistic.

**Lemma 1** *Let $U_n = \left(U_n^{[\nu]}\right)$, where $U_n^{[\nu]} = \bar{\tilde{Z}}^{[\nu]} - \mathrm{E}\left(\tilde{Z}_1^{[\nu]}\right)$ for $1 \le |\nu| \le s$, for $s = 4$. Assume that Conditions (A) hold. Then on a set having probability at least $1 - d_1/[n(\log(n))^2]$, the SRLLR statistic $R_n$ can be expanded as*

$$\begin{aligned} R_n &= \sqrt{n}T_1(U_n) + \sqrt{n}T_2(U_n) + \sqrt{n}T_3(U_n) + \frac{1}{\sqrt{n^3}}Rem_n \\ &= \sqrt{n}P(U_n) + \frac{1}{\sqrt{n^3}}Rem_n, \end{aligned} \tag{3.8}$$

*where $Rem_n$ is a remainder term that satisfies*

$$\Pr\left(\left|\frac{1}{\sqrt{n^3}}Rem_n\right| < d_2\left[\frac{\log(n)}{n}\right]^2\right) \ge 1 - \frac{d_3}{n[\log(n)]^2}.$$

*Here $T_i(U_n)$, $i = 1, 2$, and $3$, are polynomial terms of degree $i$ in the coordinates of $U_n$ and $d_1, d_2$, and $d_3$ are some constants. The main term $T_1(U_n)$ is*

$$T_1(U_n) = \frac{\partial \bar{l}_n}{\partial \theta_k}(\theta_0)\left[n\mathrm{E}\left(\frac{\partial^2 \bar{l}_n}{\partial \theta_k^2}(\theta_0)\right)\right]^{-\frac{1}{2}} \tag{3.9}$$

*expressed as a function of $U_n$. When the information matrix $I(\theta_0)$ is an identity matrix,*

$$T_1(U_n) = \frac{\partial \bar{l}_n}{\partial \theta_k}(\theta_0). \tag{3.10}$$

**Proof.** This lemma is stated by Jensen (1993) with an outlined proof. A detailed expression of (3.8) can be found in Barndorff-Nielson and Cox (1994, page 154). ■

The SRLLR statistic is used to test a scalar parameter. In general, the likelihood ratio statistics can be used to test the hypotheses $(\theta_{(k_1+1)}, \ldots, \theta_k) = (\theta_{(k_1+1)0}, \ldots, \theta_{k0})$. Let $R_{n,i}$ be the SRLLR statistic to test $\theta_i = \theta_{i0}$ under the model $(\theta_{(k_1+1)}, \ldots, \theta_k) = (\theta_{(k_1+1)0}, \ldots, \theta_{k0})$ and let $W_{n,i} = R_{n,i}^2$. Then $W_n$ can be written as

$$W_n = \sum_{i=k_1+1}^{k} W_{n,i} = \sum_{i=k_1+1}^{k} R_{n,i}^2.$$

Jensen (1993) derives and gives conditions for the existence of the Edgeworth expansion for the log likelihood ratio statistic $W_n$. We state Jensen's result in the following Lemma.

**Lemma 2 (Jensen 1993, Theorem 2.4.)** *Assume that Conditions (A) and (B1) hold for $s = 4$, that $I(\theta_0)$ in (3.6) is a $k \times k$ identity matrix, and that $A^{(12)}$ has full rank, where $A^{(12)}$ is the $m_1 \times (k - k_1)$ matrix of the first $m_1$ rows and columns $(k_1 + 1, \ldots, k)$ of $A$ in (3.7). Then there exists a polynomial $q(v)$ such that*

$$\sup_u \left|\Pr(W_n < u) - \int_0^u \left(1 + \frac{1}{n}q(v;\theta)\right) f_{k-k_1}(v)dv\right| = O\left(\frac{1}{n}\right),$$

*where $f_{k-k_1}$ is the chi-squared density with $k - k_1$ degrees of freedom and $q(v;\theta)$ is a polynomial in $v$ and a continuous function in $\theta$.*

**Proof.** Jensen(1993) uses Lemma 1 and Lemma 5 (stated in Appendix B) to establish this lemma. Jensen (1989) proves Lemma 5 and refers to Bhattacharya and Ghosh (1978) who derive the form of $q(v;\theta)$. From Remark 1.4 of Bhattacharya and Ghosh (1978), $q(v;\theta)$ depends only on the moments of $U_n$. Hence $q(v;\theta)$ is a continuous function of $\theta$ when $I(\theta)$ is a continuous function of $\theta$. Thus Condition (A1) establishes that $q(v;\theta)$ is a continuous function of $\theta$. ■

The distribution of a likelihood ratio statistic with a Bartlett adjustment can be more closely approximated by the chi-square approximation than the distribution of a likelihood ratio statistic without a Bartlett adjustment. Consider the modified statistic

$$\widetilde{W}_n = (k - k_1)\frac{W_n}{\mathrm{E}(W_n)} \tag{3.11}$$

and an expansion of $\mathrm{E}(W_n)$

$$\mathrm{E}(W_n) = (k - k_1)\left[1 + \frac{B(\theta)}{n}\right] + O\left(\frac{1}{n^2}\right). \tag{3.12}$$

Then, operationally, a Bartlett adjusted statistic $WB_n$ can be obtained by

$$WB_n = \frac{W_n}{1 + B(\tilde{\theta}^{(1)}, \theta_0^{(2)})/n}. \tag{3.13}$$

where $(\tilde{\theta}^{(1)}, \theta_0^{(2)})$ is the maximum likelihood estimate for the model parameter $\theta^{(1)}$ with the restriction $\theta^{(2)} = \theta_0^{(2)}$. The following lemma gives an asymptotic expansion for the distribution of $\widetilde{W}_n$.

**Lemma 3** *Assume the same conditions used in Lemma 2. Then there exists a polynomial $q(v;\theta)$ such that*

$$\sup_u \left|\Pr(\widetilde{W}_n < u) - \int_0^u \left(1 + \frac{1}{n}q(v;\theta)\right) f_{k-k_1}(v)dv\right| = O\left(\frac{1}{n}\right). \tag{3.14}$$

*where $f_{k-k_1}$ and $q(v;\theta)$ are the same as in the Lemma 2.*

**Proof.** Under the assumptions of Lemma 2, Jensen (1993) (Theorem 2.9) gives (3.14) with $WB_n$ instead of $\widetilde{W}_n$. Then noting that $\widetilde{W}_n = WB_n + O_p(1/n^2)$ establishes (3.14). ■

A result parallel to Lemma 2 can be obtained for the signed square root likelihood ratio statistic to test a scalar parameter $\theta_k = \theta_{k0}$. Although it is not stated in the Jensen's (1993) paper, this result is implicitly contained in it. We state the result in the following theorem and give a proof.

**Theorem 1** *Assume that Conditions (A) and (B1) hold for $s = 4$, that $I(\theta_0)$ in (3.6) is a $k \times k$ identity matrix, and that $A^{(12)}$ has full rank, where $A^{(12)}$ is the $m_1 \times (k - k_1)$ matrix of the first $m_1$ rows and columns $(k_1 + 1, \ldots, k)$ of $A$ in (3.7). Then there exists a polynomial $\xi_{s,n}(z)$ such that*

$$\sup_z \left|\Pr(R_n < z) - \int_{-\infty}^z \xi_{s,n}(v)dv\right| = O\left(\frac{1}{n}\right).$$

*where $\xi_{s,n}$ has the form*

$$\xi_{s,n}(v) = \left[1 + \sum_{j=1}^{s-2} n^{-j/2} Q_j(v)\right] \phi_{\sigma_0}(v). \tag{3.15}$$

*and where $\phi_{\sigma_0}$ is the pdf of a normal distribution with mean 0 and variance $\sigma_0$ and $Q_j$ is a polynomial in $v$.*

**Proof.** By triangular inequality we have

$$\left|\Pr(R_n < z) - \int_{-\infty}^{z} \xi_{s,n}(v)dv\right| \tag{3.16}$$

$$\leq \left|\Pr(R_n < z) - \Pr(\sqrt{n}P(U_n) < z)\right| + \left|\Pr(\sqrt{n}P(U_n) < z) - \int_{-\infty}^{z} \xi_{s,n}(v)dv\right|. \tag{3.17}$$

Note that by Lemma 1 for all $z$

$$\begin{aligned}
\Pr(R_n < z) &= \Pr\left(\sqrt{n}P(U_n) + \frac{1}{\sqrt{n^3}}Rem_n < z\right) \\
&\leq \Pr\left(\sqrt{n}P(U_n) < z + d_2\left[\frac{\log(n)}{n}\right]^2\right) + \Pr\left(\left|\frac{1}{\sqrt{n^3}}Rem_n\right| > d_2\left[\frac{\log(n)}{n}\right]^2\right) \\
&= \Pr\left(\sqrt{n}P(U_n) < z\right) + O\left(\left[\frac{\log(n)}{n}\right]^2\right) + O\left(\frac{1}{n[\log(n)]^2}\right) \qquad (3.18) \\
&= \Pr\left(\sqrt{n}P(U_n) < z\right) + O\left(\frac{1}{n[\log(n)]^2}\right).
\end{aligned}$$

where $d_2$ is defined in Lemma 1. Similarly we can have

$$\Pr\left(\sqrt{n}P(U_n) < z\right) \leq \Pr(R_n < z) + O\left(\frac{1}{n[\log(n)]^2}\right). \tag{3.19}$$

Combining (3.18) and (3.19),

$$\left|\Pr(R_n < z) - \Pr(\sqrt{n}P(U_n) < z)\right| = O\left(\frac{1}{n[\log(n)]^2}\right). \tag{3.20}$$

Because the right hand side of (3.20) implies that the approximation is $o(1/n)$ and does not depend on $z$, we have $\sup_z |\Pr(R_n < z) - \Pr(\sqrt{n}P(U_n) < z)| = o(1/n)$.

By the Lemma 5 (given in Appendix B) the supremum of the second term on the right side of (3.17) is of order $O(1/n)$. Thus the supremum of the absolute difference (3.16) is $O(1/n)$. ■

**Remark 1** *As mentioned in Jensen (1993) the theorems and lemmas in this section can be generalized for the case that $I(\theta_0)$ in (3.6) is positive definite. Let $I(\theta_0)^{-1/2}$ be a lower triangular matrix such that $I(\theta_0)^{-1/2}I(\theta_0)^{-1/2} = I(\theta_0)^{-1}$. Let $\tilde{\theta} = \theta I(\theta_0)^{-1/2}$. Then testing $\theta_{k_1+1} = \theta_{(k_1+1)0}, \dots, \theta_k = \theta_{k0}$ is the same as testing $\tilde{\theta}_{k_1+1} = \tilde{\theta}_{(k_1+1)0}, \dots, \tilde{\theta}_k = \tilde{\theta}_{k0}$, where $\tilde{\theta}_{(k_1+1)0}, \dots, \tilde{\theta}_{k0}$ are linear combinations of $\theta_{(k_1+1)0}, \dots, \theta_{k0}$. The condition that $A^{(12)}$ has full rank is then replaced by Condition (B3).*

### 3.2.2 Bootstrap Statistics

Suppose that $S$ is a statistic. The parametric bootstrap version $S^*$ of $S$ is the same function but evaluated at data ("bootstrap samples") simulated using an estimate $\widehat{\theta}$ instead of the unknown $\theta$.

Denote the cumulants of $U_n$ of order $r$ by $\lambda_{j_1\dots j_r}$. That is,

$$\lambda_{j_1\dots j_r}(\theta) = \frac{1}{i^r}\frac{\partial^m \log[\mathrm{E}_\theta(e^{itU_n})]}{\partial t_{j_1}\dots\partial t_{j_r}}\bigg|_{t=0} \tag{3.21}$$

where $j_1, \dots, j_r \in \{1, \dots, m_0\}$ and $m_0$ is the dimension of $U_n$. Define

$$P_{j_1\dots j_r}(\theta) = \frac{\partial^r P(\mu + \mathrm{E}_\theta(U_n))}{\partial \mu_{j_1}\dots\partial \mu_{j_r}}\bigg|_{\mu=0}, \tag{3.22}$$

where $P$ and $U_n$ are defined in Lemma 1.

The following lemma provides the convergence rate of the difference between some statistics and their bootstrap versions. It will be used in our main theorems. Note that the bootstrap version of $\lambda_{j_1\dots j_r}$ and $P_{j_1\dots j_r}$ are $\lambda^*_{j_1\dots j_r} = \lambda_{j_1\dots j_r}(\widehat{\theta})$ and $P^*_{j_1\dots j_r} = P_{j_1\dots j_r}(\widehat{\theta})$, respectively.

**Lemma 4** *Assume that Conditions (A) and (B) hold for some specified $s \geq 4$ and that $g_1$ is a continuous differentiable function. Then*

$$\lambda_{j_1\dots j_r} - \lambda^*_{j_1\dots j_r} = O_p\left(\frac{1}{\sqrt{n}}\right), \quad P_{j_1\dots j_r} - P^*_{j_1\dots j_r} = O_p\left(\frac{1}{\sqrt{n}}\right),$$

$$g_1(\lambda_{j_1\dots j_r}) - g_1(\lambda^*_{j_1\dots j_r}) = O_p\left(\frac{1}{\sqrt{n}}\right), \quad g_1(P_{j_1\dots j_r}) - g_1(P^*_{j_1\dots j_r}) = O_p\left(\frac{1}{\sqrt{n}}\right).$$

**Proof.** Under the stated conditions, $\sqrt{n}(\widehat{\theta} - \theta)$ has a limiting normal distribution (Jensen (1993), Theorem 2.1). So we have $\widehat{\theta} - \theta = O_p(1/\sqrt{n})$. By Condition (A1), $\lambda_{j_1\dots j_r}$ is a continuous differentiable function of $\theta$ and $\lambda^*_{j_1\dots j_r}$ is a continuous differentiable function of $\widehat{\theta}$. Note that $\lambda^*_{j_1\ldots j_r} = \lambda_{j_1\ldots j_r}(\widehat{\theta})$. The delta method shows that $\lambda_{j_1\ldots j_r}(\theta) - \lambda_{j_1\ldots j_r}(\widehat{\theta}) = O_p(1/\sqrt{n})$. Also $g_1[\lambda_{j_1\ldots j_r}(\theta)] - g_1[\lambda_{j_1\ldots j_r}(\widehat{\theta})] = O_p(1/\sqrt{n})$.

Note that $\mathrm{E}(U^*_n) = \mathrm{E}\left[\dot{Z} - E(\dot{Z}_1)\right] = 0$ and $P$ is a polynomial. From (3.22) we see that $P_{j_1\ldots j_r}/r$ are coefficients of $T_r(\cdot)$ and these coefficients are continuous differentiable functions of $\theta$. Also $P^*_{j_1\ldots j_r} = P_{j_1\ldots j_r}(\widehat{\theta})$. Hence we have that $P_{j_1\ldots j_r} - P^*_{j_1\ldots j_r} = O_p(1/\sqrt{n})$ and $g_1(P_{j_1\ldots j_r}) - g_1(P^*_{j_1\ldots j_r}) = O_p(1/\sqrt{n})$. ■

The following Theorem establishes the result that the distribution of the SRLLR statistic can be approximated to order $O(1/n)$ by its bootstrap distribution.

**Theorem 2** *Assume that Conditions (A) and (B) hold for some specified $s \geq 4$. Then*

$$\sup_z |\Pr(R_n \leq z) - \Pr(R^*_n \leq z)| = O_p\left(\frac{1}{n}\right).$$

*Proof.* By the Lemma 5 we have the result of the Corollary B in Appendix B. That is

$$\Pr(\sqrt{n}P(U^*_n) < z) = \Phi_{\sigma^2}(z) - \frac{1}{\sqrt{n}}a_1(z)\phi_{\sigma^2}(z) + O\left(\frac{1}{n}\right).$$

where $\Phi_{\sigma^2}$ and $\phi_{\sigma^2}$ denote the cdf and pdf of a normal distribution with mean 0 and variance $\sigma^2$, $\sigma^2 = \sum_{j,k=1,m_0} \lambda_{jk}P_jP_k$, $a_1(z)\phi_{\sigma^2}(z)$ is bounded over $z$, and $a_1$ is a continuous differentiable function of $\lambda_{j_1\ldots j_r}$ and $P_{j_1\ldots j_r}$. By the proof in Theorem 1 we have $\sup_z |\Pr(R_n < z) - \Pr(\sqrt{n}P(U^*_n) < z)| = O(1/n)$. This implies that for all z.

$$\Pr(R_n < z) = \Phi_{\sigma^2}(z) - \frac{1}{\sqrt{n}}a_1(z)\phi_{\sigma^2}(z) + O\left(\frac{1}{n}\right). \tag{3.23}$$

By using Lemma 1 together with $\lambda_{ij}$ defined in (3.21) and $P_j$, $P_k$ defined in (3.22), we have $\lambda_{kk} = \mathrm{Var}(U^{*[e_k]}_n)$, where $U^{*[e_k]}_n$ is the $k$th coordinate of $U^*_n$, and $P_j = 0$, $j \neq k$, and $P_k = \left[\mathrm{E}\left(\frac{\partial^2 \bar{l}_n}{\partial\theta_k^2}(\theta_0)\right)\right]^{-\frac{1}{2}}$. Thus

$$\begin{aligned}\sigma^2 &= \sum_{j,k=1,m_0} \lambda_{jk}P_jP_k = \lambda_{kk}P_kP_k = \mathrm{Var}(U^{*[e_k]}_n)\left[\mathrm{E}\left(\frac{\partial^2 \bar{l}_n}{\partial\theta_k^2}(\theta_0)\right)\right]^{-1} \\ &= \mathrm{Var}\left[\frac{\partial \bar{l}_n}{\partial\theta_k}(\theta_0)\right]\left[\mathrm{E}\left(\frac{\partial^2 \bar{l}_n}{\partial\theta_k^2}(\theta_0)\right)\right]^{-1} = 1.\end{aligned} \tag{3.24}$$

Then (3.23) becomes

$$\Pr(R_n < z) = \Phi_1(z) - \frac{1}{\sqrt{n}}a_1\phi_1(z) + O\left(\frac{1}{n}\right). \tag{3.25}$$

Assume that Conditions (A) hold in a parametric bootstrap scheme and suppose that data are sampled from the distribution with true parameter $\widehat{\theta}$. By Lemma 1 and definitions in (3.21) and (3.22), we have $\lambda^*_{kk} = \mathrm{Var}(l_n^{*[e_k]*})$, $P^*_j = 0$, $j \neq k$, and $P^*_k = \left[\mathrm{E}\left(\frac{\partial^2 l^*_n}{\partial \theta_k^2}(\widehat{\theta})\right)\right]^{-\frac{1}{2}}$. Thus

$$\sigma^{2*} = \sum_{j,k=1,m_0} \lambda^*_{jk} P^*_j P^*_k = \lambda^*_{kk} P^*_k P^*_k = \mathrm{Var}(l_n^{*[e_k]*}) \left[\mathrm{E}\left(\frac{\partial^2 \bar{l}^*_n}{\partial \theta_k^2}(\widehat{\theta})\right)\right]^{-1}$$

$$= \mathrm{Var}\left[\frac{\partial \bar{l}^*_n}{\partial \theta_k}(\widehat{\theta})\right] \left[\mathrm{E}\left(\frac{\partial^2 \bar{l}^*_n}{\partial \theta_k^2}(\widehat{\theta})\right)\right]^{-1} = 1. \tag{3.26}$$

Hence the bootstrap version of (3.23) is

$$\Pr(R^*_n < z) = \Phi_1(z) - \frac{1}{\sqrt{n}} a^*_1(z)\phi_1(z) + O_p\left(\frac{1}{n}\right). \tag{3.27}$$

By Lemma 4 we have

$$a_1 - a^*_1 = O_p\left(\frac{1}{\sqrt{n}}\right).$$

The difference between Equation (3.25) and Equation (3.27) is $O_p(1/n)$. This implies

$$|\Pr(R_n \le z) - \Pr(R^*_n \le z)| = O_p\left(\frac{1}{n}\right). \tag{3.28}$$

Because $a_1(z)\phi_1(z)$ and $a^*_1(z)\phi_1(z)$ are bounded over $z$,

$$\sup_z |a_1(z)\phi_1(z)/\sqrt{n} - a^*_1(z)\phi_1(z)/\sqrt{n}| = O_p(1/n).$$

Thus we have the supremum over $z$ of (3.28) that gives the needed result.■

The following shows that a similar result can be obtained for the log likelihood ratio statistic.

**Theorem 3** *Assume that Conditions (A) and (B) hold for some specified $s \ge 4$. Then for all $z \ge 0$ we have*

$$\sup_z |\Pr(W_n \le z) - \Pr(W^*_n \le z)| = O_p\left(\frac{1}{n}\right).$$

*Proof.* Note that

$$\begin{aligned}
&|\Pr(W_n \le z) - \Pr(W^*_n \le z)| \\
&\le \left|\Pr(W_n < u) - \int_0^u \left(1 + \frac{1}{n} q(v;\theta)\right) f_{k-k_1}(v)dv\right| \\
&+ \left|\int_0^u \left(1 + \frac{1}{n} q(v;\theta)\right) f_{k-k_1}(v)dv - \int_0^u \left(1 + \frac{1}{n} q^*(v;\widehat{\theta})\right) f_{k-k_1}(v)dv\right| \qquad (3.29) \\
&+ \left|\Pr(W^*_n < u) - \int_0^u \left(1 + \frac{1}{n} q^*(v;\widehat{\theta})\right) f_{k-k_1}(v)dv\right|.
\end{aligned}$$

Because $q(v:\theta)$ in Lemma 2 is a continuous function of $\theta$, $q(x:\theta) - q(x:\widehat{\theta}) = O_p(1/\sqrt{n})$. Also because $q(v:\theta)$ is a polynomial in $v$ and $f_{k-k_1}$ is the density of the chi-square distribution, $\int_{-\infty}^{\infty} |q(v:\theta)| f_{k-k_1}(v)dv$ is finite. Because

$$\int_{-\infty}^{u} q(v:\theta) f_{k-k_1}(v)dv \leq \int_{-\infty}^{\infty} |q(v:\theta)| f_{k-k_1}(v)dv.$$

$\int_{-\infty}^{u} q(v:\theta) f_{k-k_1}(v)dv$ is bounded over $u$. Thus

$$\sup_u \left| \int_0^u q(v:\theta) f_{k-k_1}(v)dv - \int_0^u q^*(v:\widehat{\theta}) f_{k-k_1}(v)dv \right| = O_p(1) \tag{3.30}$$

which implies that

$$\sup_u \left| \int_0^u \frac{1}{n} q(v:\theta) f_{k-k_1}(v)dv - \int_0^u \frac{1}{n} q^*(v:\widehat{\theta}) f_{k-k_1}(v)dv \right| = O_p\left(\frac{1}{n}\right). \tag{3.31}$$

By (3.31) and by using Lemma 2 for the first and third term on the right hand side of (3.29), we have the result. ■

The following shows that a similar result can be obtained for log likelihood ratio statistics with a Bartlett correction.

**Theorem 4** *Assume that Conditions (A) and (B) hold for some specified $s \geq 4$. Then for all $z \geq 0$ we have*

$$\sup_z \left| \Pr(\widetilde{W}_n \leq z) - \Pr(\widetilde{W}_n^* \leq z) \right| = O_p\left(\frac{1}{n}\right).$$

*Proof.* Using the result of Lemma 3 and arguments in Theorem 3 will establish the result in this theorem. ■

## 3.3 Examples

### 3.3.1 Confidence Interval (Bound) Procedures Based on Likelihood Ratio Statistics

Jensen (1987) gives an example with fixed censoring time under the one-parameter exponential model. He uses simulation to show that with 15 failures and a proportion failing equal to .75, the large sample approximation using likelihood ratio statistics (for two-sided intervals) and its signed square root form (for one-sided bounds), based on the theory in the previous section, provides accurate coverage probabilities. But when censoring is heavy (e.g., the proportion failing is equal to .1 and sample size is 20) the upper confidence bound constructed from the signed square root likelihood ratio statistic is very conservative. This is consistent with the results in Jeng and Meeker (1998). Moreover, Jeng and Meeker (1998) show that in the two parameter Weibull model, with heavy time censoring (proportion failing $\leq .1$), the approximate procedure based on likelihood ratio statistic for constructing the one-sided confidence bounds is not accurate when the number of failure is less than 20.

Jensen (1989) shows that the exponential lifetime model with random censoring satisfies Conditions (A) and (B). Hence Lemma 2 in Section 3.2 can provide a second order accurate procedure for constructing two-sided confidence intervals. Jensen (1993) presents an application using the logistic regression model. His numerical simulation results suggest that the coverage probability of the procedure using likelihood ratio statistic with a Bartlett correction for constructing two-sided confidence intervals has fourth order accuracy $[O(1/n^2)]$ when the sample size is more than 20.

### 3.3.2 The Two Parameter Weibull Distribution Model

To explore the finite sample performance of the asymptotic results in Section 3.2 we conduct a simulation study using the two-parameter Weibull distribution model with complete and Type I censored data.

#### 3.3.2.1 Regularity Conditions

We describe the formulation for the general location-scale distribution model. The logarithm of a Weibull random variable has a smallest extreme value distribution which belongs to the location-scale family. Suppose that the continuous random variable $X = \log(T)$ has density $\phi[(x-\mu)/\sigma]/\sigma$ and cdf $\Phi[(x-\mu)/\sigma]$, where $(\mu, \sigma) = \theta$ is the unknown parameter in an open set $\Theta \subset \mathbb{R}^2$. Let $t_c$ denote the censoring time and define $\delta = 1$ for a failure and $\delta = 0$ for a censored observation. The observations are $x_1 = \log(t_1), \ldots, x_n = \log(t_n)$. Let $x_c = \log(t_c)$. The log likelihood of an observation $x_i$ is

$$\begin{aligned} l(x_i; \theta) = \delta_i & \left\{ -\log(\sigma) + \log\left[\phi\left(\frac{x_i-\mu}{\sigma}\right)\right]\right\} \\ & + (1-\delta_i)\log\left[1-\Phi\left(\frac{x_c-\mu}{\sigma}\right)\right]. \end{aligned} \tag{3.32}$$

We could be interested in the location or the scale parameter or in a particular quantile or other function of the parameters. We do the development for estimating a particular quantile. Other functions of the parameters can be obtained analogously. Let $x_p$ be the $p$ quantile of the distribution $\Phi[(x-\mu)/\sigma]$, and $u_p = \Phi^{-1}(p)$. Then $x_p = \mu + u_p\sigma$ and $t_p = \exp(x_p)$ is the $p$ quantile of the Weibull distribution. The confidence intervals (bounds) for $t_p$ can be obtained by taking the exp transformation of the confidence intervals (bounds) for $x_p$. The likelihood in (3.32) can be rewritten as a function of $(\sigma, x_p)$

$$\begin{aligned} l(x_i;(\sigma,x_p)) = \delta_i \left\{ -\log(\sigma) + \log\left[\phi\left(\frac{x_i - x_p}{\sigma} - u_p\right)\right]\right\} + \\ (1-\delta_i)\log\left[1 - \Phi\left(\frac{x_c - x_p}{\sigma} - u_p\right)\right]. \end{aligned} \tag{3.33}$$

With $l$ smooth enough and $\phi$ having light tails, it can be shown that Conditions (A) stated in section 3.2 are satisfied. See Appendix B for detail. Then for $|\nu| \leq 4$,

$$\begin{aligned} Z_i = \Bigg( & \frac{\partial l(X_i;(\sigma,x_p))}{\partial x_p}, \frac{\partial l(X_i;(\sigma,x_p))}{\partial \sigma}, \frac{\partial^2 l(X_i;(\sigma,x_p))}{\partial x_p^2}, \frac{\partial^2 l(X_i;(\sigma,x_p))}{\partial x_p \partial \sigma}, \\ & \dots, \frac{\partial^4 l(X_i;(\sigma,x_p))}{\partial \sigma^4}\Bigg), \end{aligned} \tag{3.34}$$

where $Z_i$ is a 14 dimensional vector. Transform $Z_i$ into a $m_0 = m_1 + m_2$ dimensional vector $\tilde{Z}_i$ with linearly independent coordinates for which the first $m_1$ ordinates are continuous and last $m_2$ coordinates are discrete. The form of $\tilde{Z}_i$ depends on the distribution of the observations. For the SEV, normal, and logistic distributions $\tilde{Z}_i$ is shown in Appendix B. Note that $\delta_i$ is the only discrete part of $Z_i$, so it is the only discrete part of $\tilde{Z}_i$. By Proposition 1, Condiction (B1) is satisfied here.

The first two elements of $Z_i$ are linearly independent when data come from the SEV, normal or logistic distribution (see Appendix B). The first two elements of the first two columns of $A^{(11)}$ are $(1,0)$ and $(c_1, c_2)$ respectively, where $c_1, c_2$ are non-zero constants (that could depend on the parameters), hence $A^{(11)}$ has full rank 2. For the SEV, normal, and logistic distributions $(c_1, c_2)$ is just $(0,1)$. So the Condiction (B2) holds.

Because the first $m_1$ rows of $A^{(1)}$ gives $A^{(11)}$ as described above and $I(\theta_0)^{-1/2}$ is a lower triangular positive definite matrix, $(A^{(1)}I(\theta_0)^{-1/2})^{(11)}$ is a $m_1$ dimentional vector which has rank 1. Thus Condiction (B3) holds. The theorems in Section 3.2 tell us that the procedure based on the bootstrap log likelihood ratio statistic or its corresponding signed square root procedures can be used to construct two-sided (one-sided) confidence intervals (bounds) that are second order accurate.

### 3.3.3 Simulation Study

#### 3.3.3.1 Simulation Set Up

Let $T$ be a random variable having a Weibull distribution, then $X = \log(T)$ has a smallest extreme value (SEV) distribution with density $\phi_{SEV}(z)/\sigma$ and cdf $\Phi_{SEV}(z)$, where $\phi_{SEV}(z) = \exp[-z - \exp(z)]$, $\Phi_{SEV}(z) = 1 - \exp[-\exp(z)]$ and $z = (x - \mu)/\sigma$, $-\infty < x < \infty, -\infty < \mu < \infty, \sigma > 0$. Our simulation was designed to study the following experimental factors:

- $p_f$: the expected proportion failing by the censoring time.
- $\mathrm{E}(r) = np_f$: the expected number of failures before the censoring time.

We used 5000 Monte Carlo samples for each $p_f$ and $\mathrm{E}(r)$ combination. The number of bootstrap replications was $B = 10000$. The levels of the experimental factors used were $p_f = .01, .1, .3, .5, .9, 1$ and $\mathrm{E}(r) = 3, 5, 7, 10, 15$ and 20. For each Monte Carlo sample we obtained the ML estimates of the scale parameter and the quantiles $\log(t_p)$, $p = .01, .05, .1, .3, .5, .632$ and $.9$, where $\mu \cong \log(t_{.632})$. The one-sided $100(1-\alpha)\%$ confidence bounds (CBs) were calculated for $\alpha$ =.025 and .05. Hence the 90% and 95% two-sided CIs can be obtained by combining the upper and lower CBs. Without loss of generality, we sampled from an SEV distribution with $\mu = 0$ and $\sigma = 1$.

Because the number of failures before the censoring time $t_c$ is random, it is possible to have as few as $r = 0$ or 1 failures in the simulation, especially when $\mathrm{E}(r)$ is small. With $r = 0$, ML estimates do not exist. With $r = 1$, LR intervals may not exist. Therefore, we calculate the results conditionally on the cases with $r > 1$.

Let $1-\alpha$ be the nominal coverage probability (CP) of a procedure for constructing a confidence interval, and let $1-\hat{\alpha}$ denote the corresponding Monte Carlo evaluation of the actual coverage probability $1-\alpha'$. The standard error of $\hat{\alpha}$ is approximately $\mathrm{se}(1-\hat{\alpha}) = [\alpha'(1-\alpha')/n_s]^{1/2}$, where $n_s$ is the number of Monte Carlo simulation trials. For a 95% confidence interval from 5000 simulations the standard error of the CP estimation is $[.05(1-.95)/5000]^{1/2} = .0031$ if the procedure is correct. The Monte Carlo error is approximately $\pm 1\%$. We say the procedure or the method for the 95% confidence region is adequate if the CP is within $\pm 1\%$ error of the normnal CP.

The modified signed square root LLR statistic is presented by Barndorff-Nelson (1986, 1991) and is asymptotically standard normal distributed with error of order $O(1/n^{3/2})$ when there is no censoring. It is a modification of the SRLLR methods.

Table 3.1 Abbreviations of the methods in simulation study

| | |
|---|---|
| LLR | Log likelihood ratio |
| LLRB | Log likelihood ratio Bartlett corrected |
| MSRLLR | Modified signed square root LLR |
| PTBT | Parametric transformed bootstrap-$t$ |
| PBBCA | Parametric bootstrap bias-corrected accelerated |
| PBSRLLR | Parametric bootstrap signed squared root LLR |
| PBMSRLLR | Parametric bootstrap MSRLLR |

We expected the PBMSRLLR will have similar or better performance as the PBSR-LLR methods. Detailed descriptions of the methods for constructing these confidence intervals are given in Appendix B.

#### 3.3.3.2 Simulation Results

This section presents some of the most interesting and useful results from our simulation. Figure 3.1 shows the coverage probability of the procedures for the one-sided approximate 95% CBs for the parameter $\sigma$ from the seven methods for five different proportion failing values. Figure 3.2 is the same type of graph as Figure 3.1 for $t_{.1}$, the .1 quantile of Weibull distribution. Figure 3.3 shows CPs of these procedures when $p_f = .5$ for different quantiles. Figures 3.4 to 3.7 present a closer comparison of CP for methods and parameters. Figure 3.8 and Figure 3.9 show the coverage probability of these procedures for 90% two-sided confidence intervals. We summarize the simulation results briefly as follows:

- The LLR method with a Bartlett correction does not improve the coverage probability of the procedure for one-sided confidence bounds. For one-sided confidence bounds, the LLR and LLRB methods are adequate when the expected number of failures $\geq 20$. For two-sided confidence intervals, the LLR method is adequate when the expected number of failures is more than 15 and the LLR method with a Bartlett correction is very accurate even for an expected number of failures as small as 7.

- When there is no censoring, the MSRLLR method is an accurate procedure for one-sided confidence bounds, even with the expected number of failures as small as 5. For Type I censoring, the coverage probability of the MSRLLR method depends on the proportion failing, the expected number of failures, and the parameters of interest. Generally, the MSRLLR method for the one-sided confidence bounds and two-sided confidence intervals is adequate when the expected number of failures exceeds 20.

- The bootstrap-$t$ method is an accurate procedure for the scale parameter. When the quantity of interest is the $p$ quantile, where $p$ is close to the proportion failing, the one-sided lower confidence bound procedure is anti-conservative. The bootstrap-$t$ method gives accurate coverage probabilities for all functions of the parameters when the number of failures exceeds 20.

- The $BC_a$ method for both one-sided confidence bounds and two-sided confidence intervals is adequate when the number of failures exceeds 20.

- The PBSRLLR method for the one-sided confidence bounds and two-sided confidence intervals is adequate except when the number of failures is less than 15 and the quantity of interest is the $p$ quantile where $p$ is close to the proportion failing.

- Among these seven methods, the PBMSRLLR method is most accurate for one-sided confidence bounds. When the number of failures is less than 10 and the quantity of interest is the $p$ quantile where $p$ is close to the proportion failing, the PBMSRLLR method for the one-sided confidence bound is less accurate. Generally the PBMSRLLR method is adequate when the number of failures is 10 or more. For two-sided confidence intervals, the PBMSRLLR method is adequate when the expected number of failures exceeds 7.

For Type I censored data, we can draw the following conclusion. If our interest is in constructing one-sided confidence bounds, the PBSRLLR and PBMSRLLR methods provide better coverage probability with a small expected number of failures (like 10). For two-sided confidence intervals, the PBSRLLR, PBMSRLLR and LLRB methods provide accurate procedures. The LLRB and PBMSRLLR methods give more accurate results even when the expected number of failures is as small as 7. The two-sided confidence interval from the PBSRLLR or the PBMSRLLR method is more symmetric than that from other methods in the sense that the confidence level of one side of the interval is close to the confidence level of the other side of the interval.

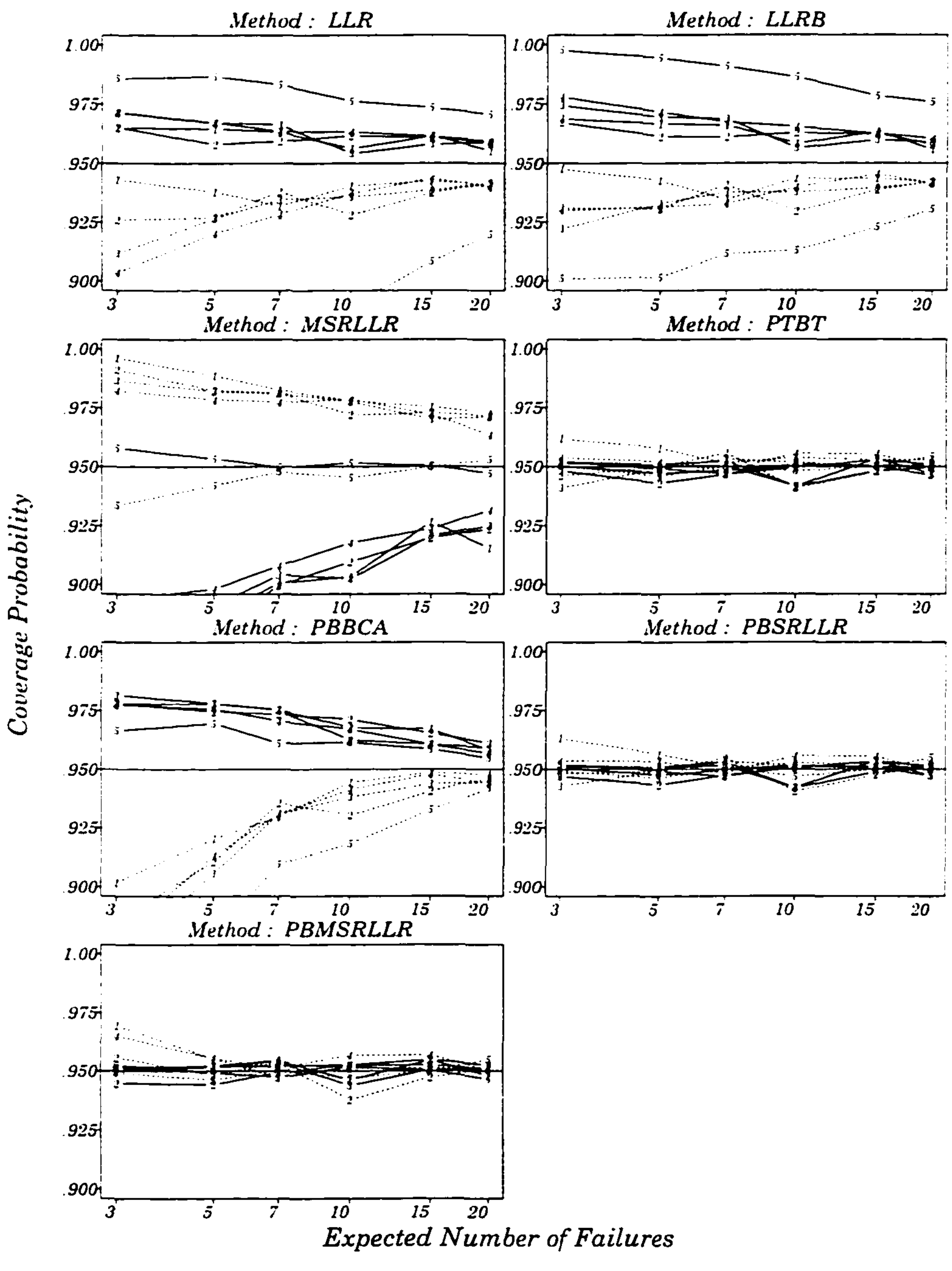



Figure 3.1 Coverage probability versus expected number of failures plot for one-sided approximate 95% CI procedures for parameter $\sigma$. The numbers (1, 2, 3, 4, 5) in the lines of each plot correspond to $p_f$'s (.01, .1, .3, .5, 1). Dotted and solid lines correspond to upper and lower bounds, respectively.

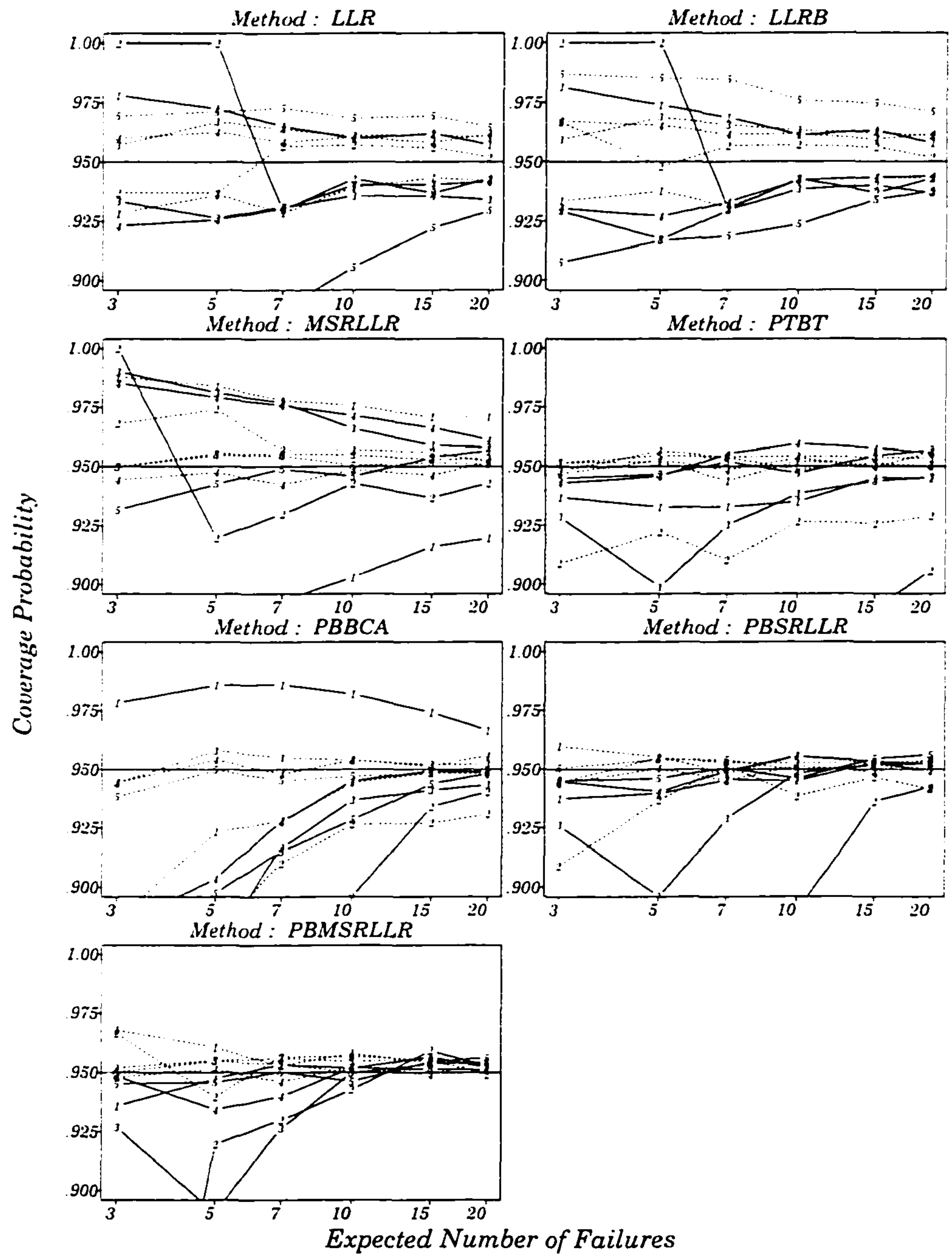



Figure 3.2 Coverage probability versus expected number of failures plot for one-sided approximate 95% CI procedures for parameter $t_{.1}$. The numbers (1, 2, 3, 4, 5) in the lines of each plot correspond to $p_f$'s (.01, .1, .3, .5, 1). Dotted and solid lines correspond to upper and lower bounds, respectively.

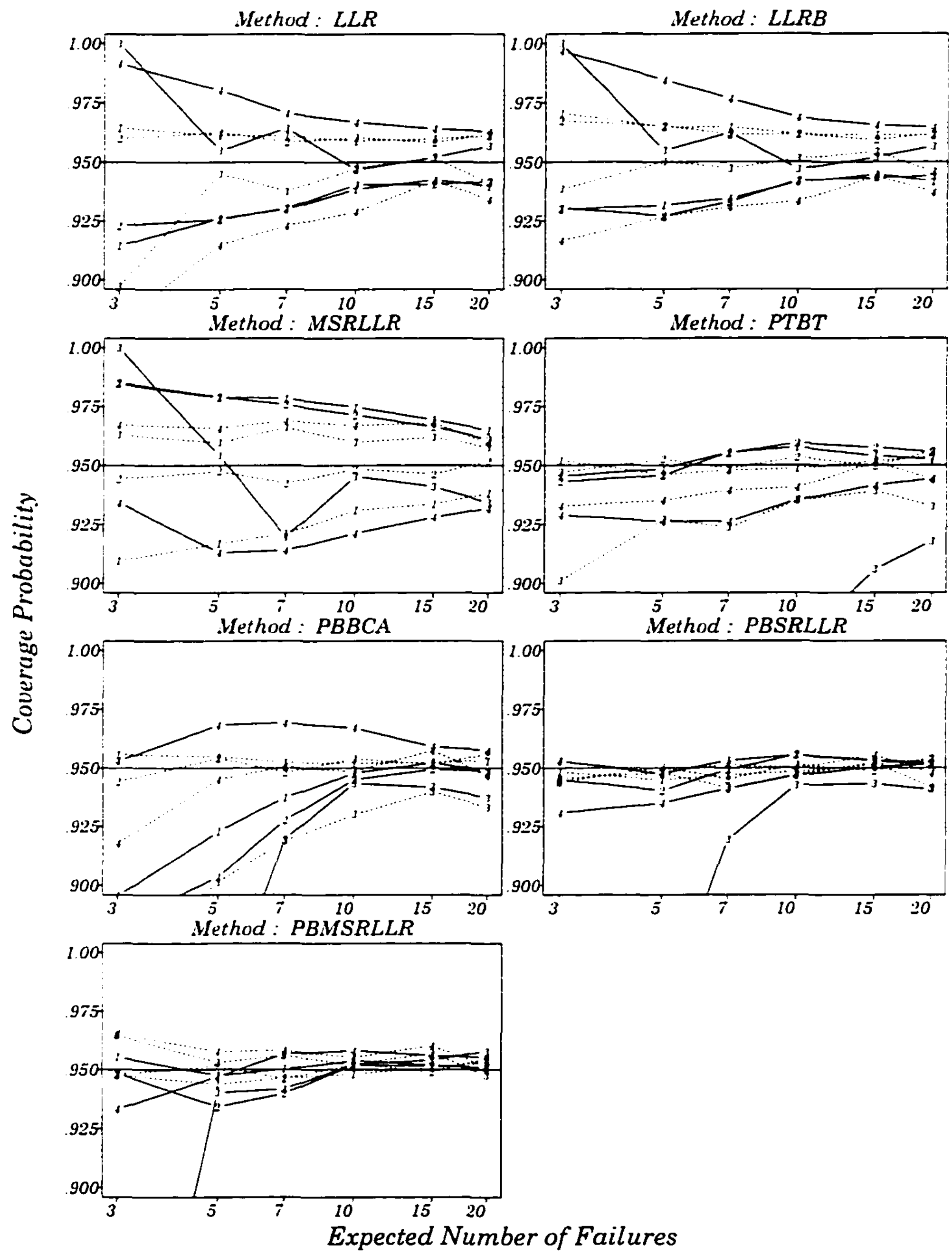



Figure 3.3 Coverage probability versus expected number of failures plot for one-sided approximate 95% CI procedures for proportion failing $p_f = .5$. The numbers (1, 2, 3, 4, 5) in the lines of each plot correspond to $t_p$'s, $p = (.01, .1, .5, .632, .9)$. Dotted and solid lines correspond to upper and lower bounds, respectively.

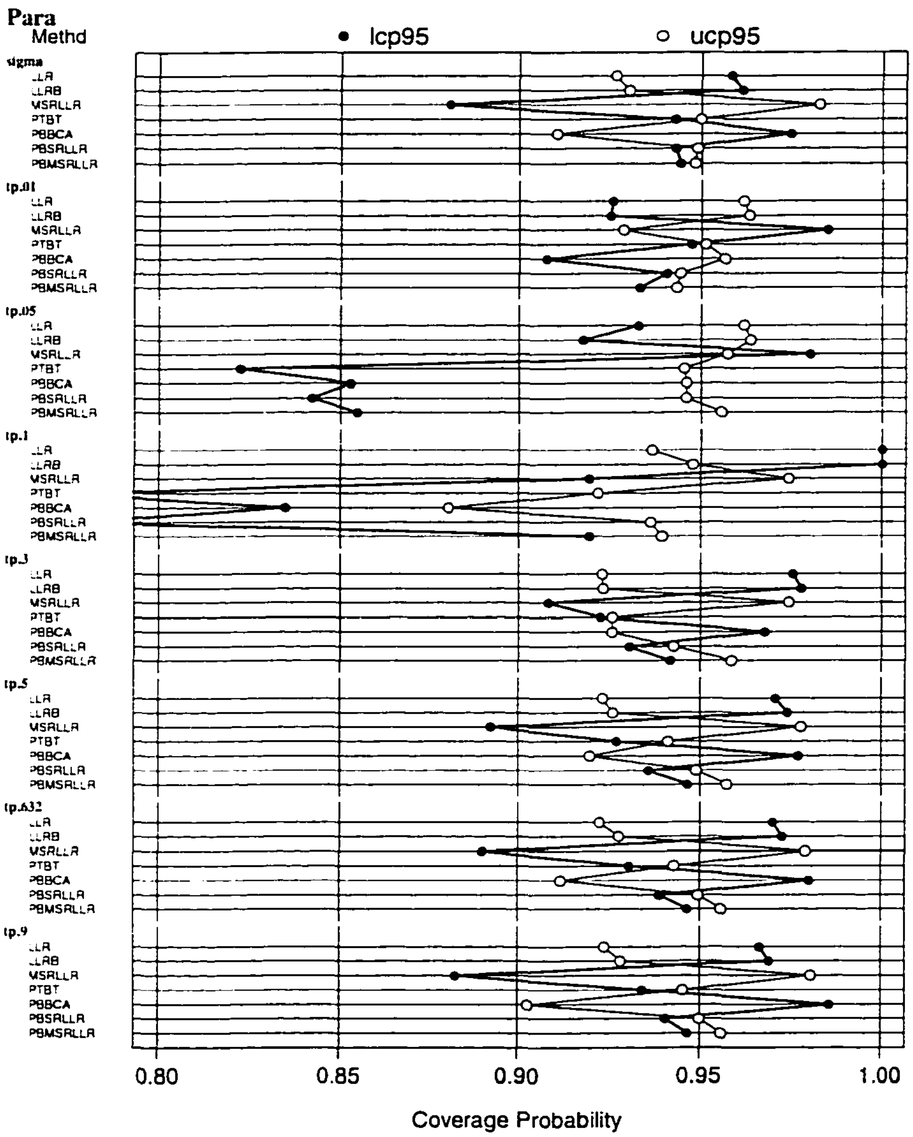



Figure 3.4 Coverage probability plot for approximate 95% one-sided confidence interval procedures in the case $E(r) = 5$ and $p_f = .1$.

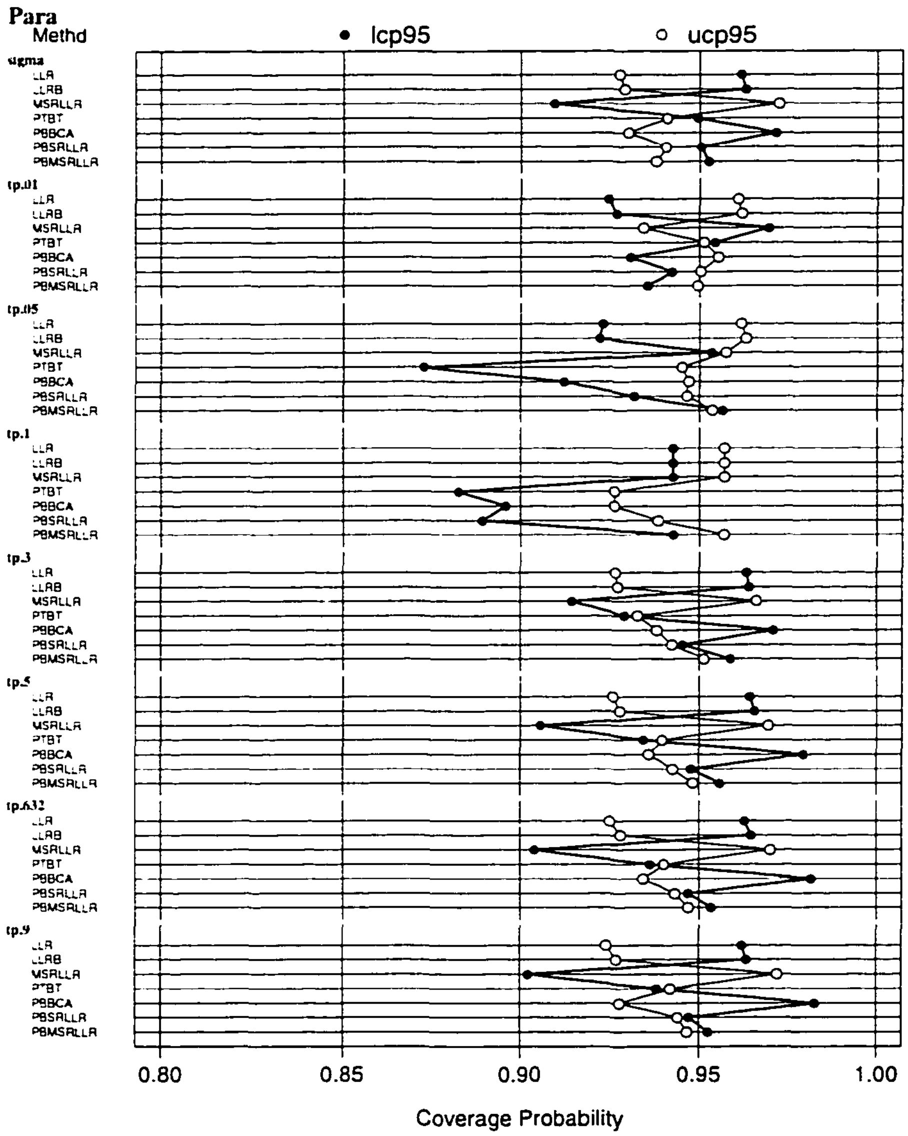



Figure 3.5 Coverage probability plot for approximate 95% one-sided confidence interval procedures in the case $E(r) = 10$ and $p_f = .1$.

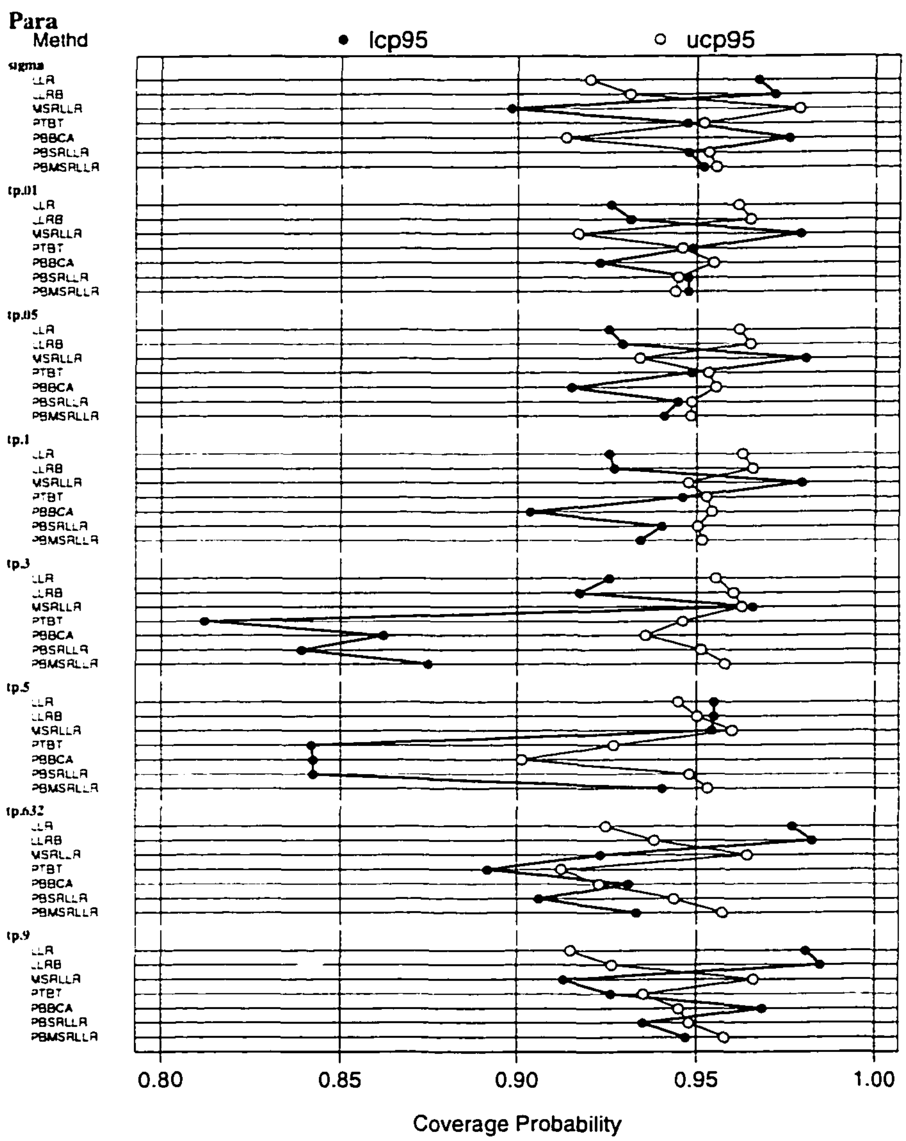



Figure 3.6 Coverage probability plot for approximate 95% one-sided confidence interval procedures in the case $\mathrm{E}(r) = 5$ and $p_f = .5$.

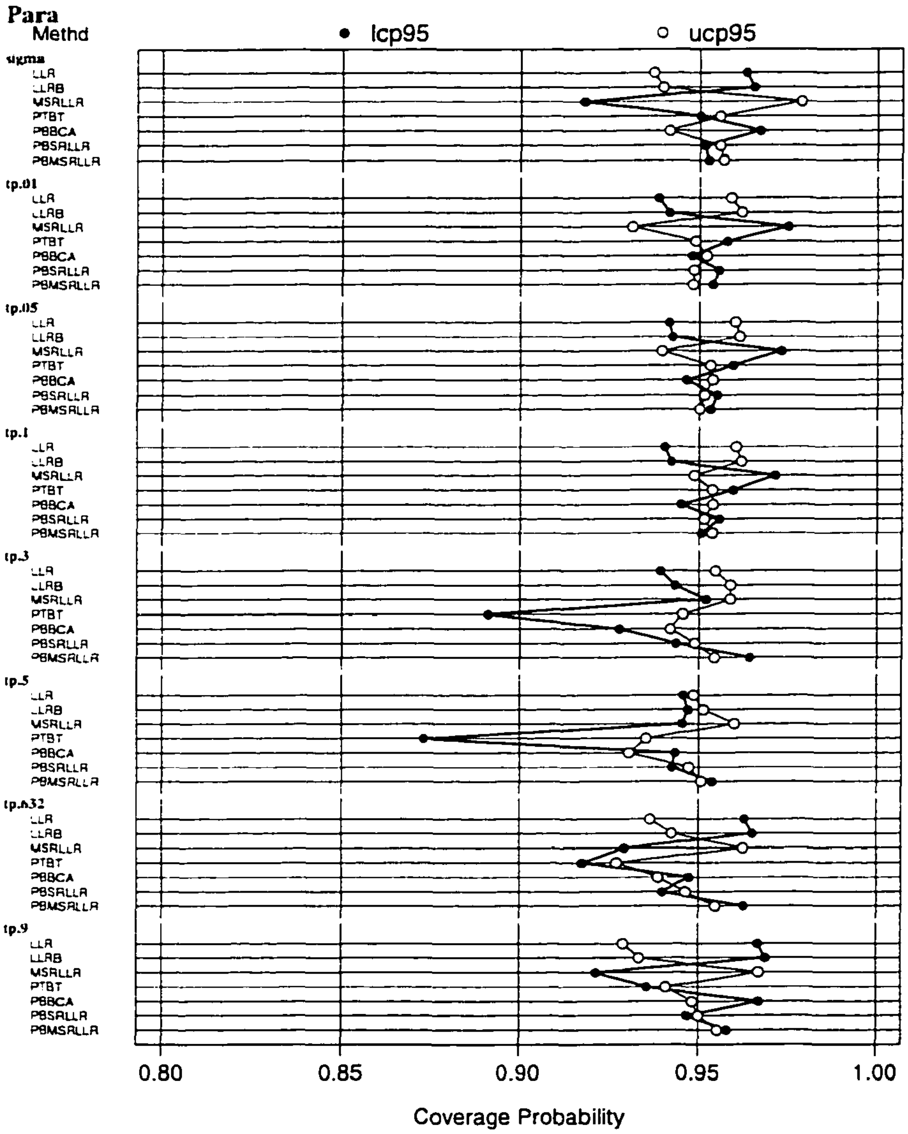



Figure 3.7 Coverage probability plot for approximate 95% one-sided confidence interval procedures in the case E$(r)$ = 10 and $p_f$ = .5.

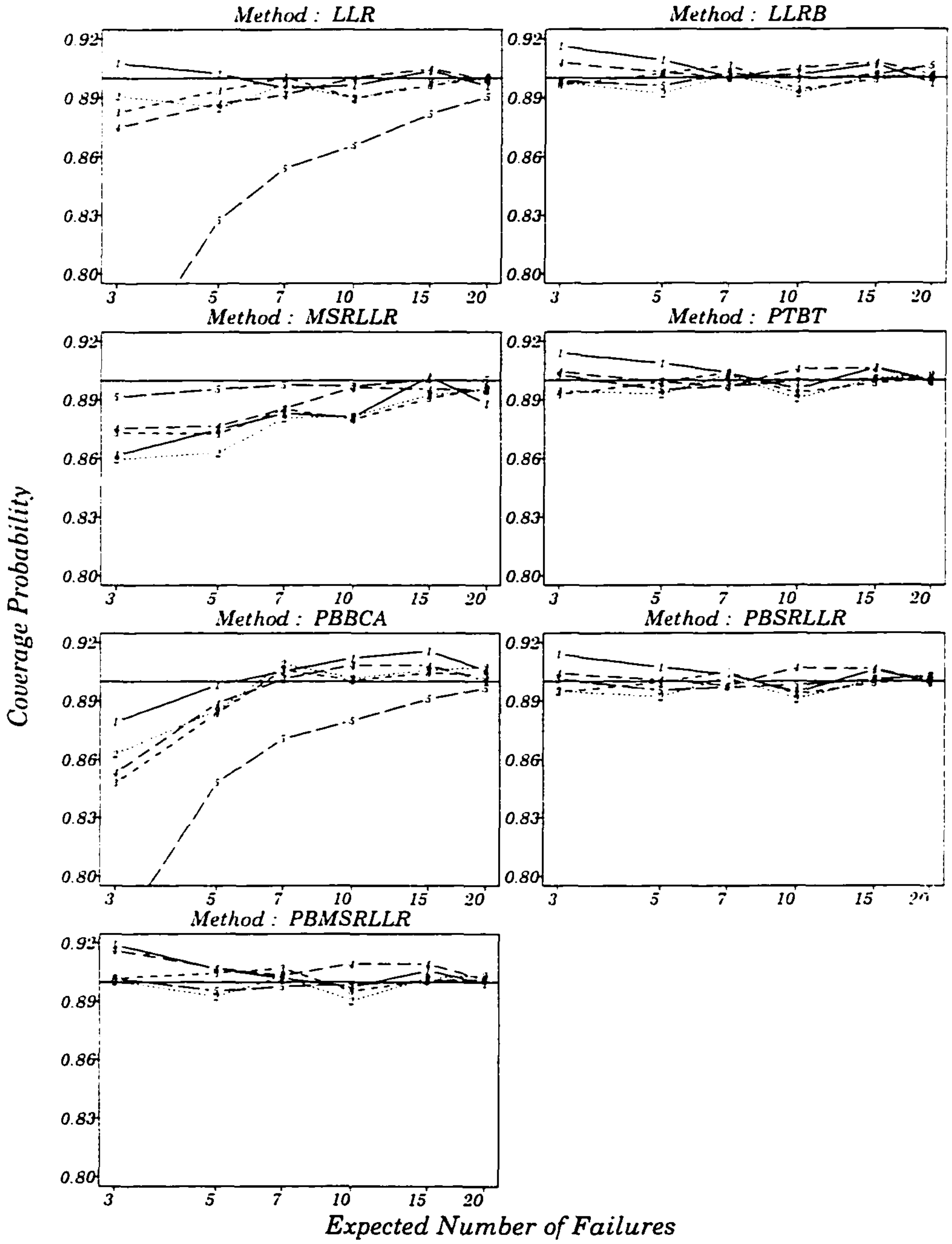



Figure 3.8 Coverage probability versus expected number of failures plot for two-sided approximate 90% CI procedures for parameter $\sigma$. The numbers (1, 2, 3, 4, 5) in the lines of each plot correspond to $p_f$'s (.01, .1, .3, .5, 1). Dotted and solid lines correspond to upper and lower bounds, respectively.

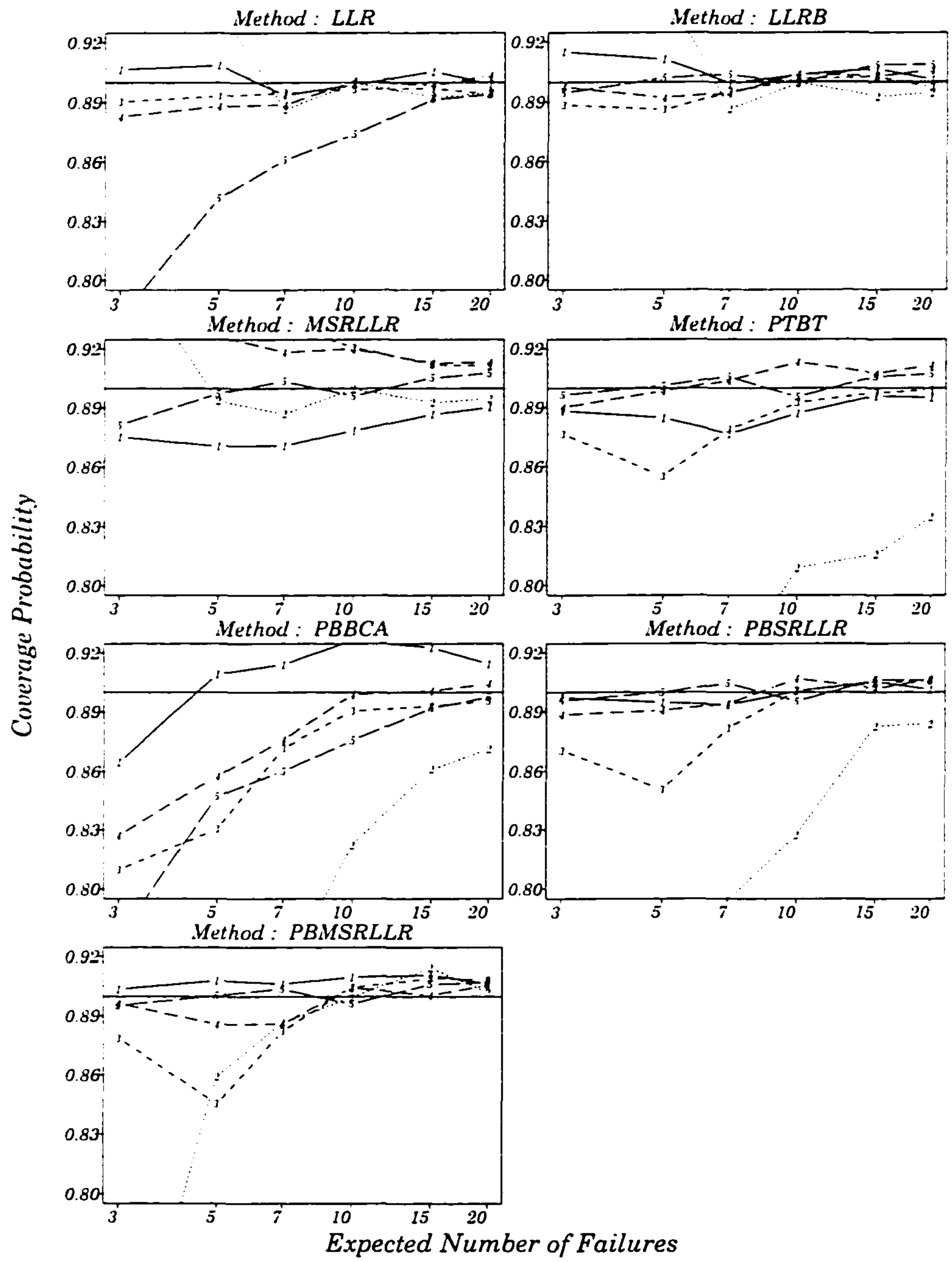



Figure 3.9 Coverage probability versus expected number of failures plot for two-sided approximate 90% CI procedures for parameter $t_{.1}$. The numbers (1, 2, 3, 4, 5) in the lines of each plot correspond to $p_f$'s (.01, .1, .3, .5, 1). Dotted and solid lines correspond to upper and lower bounds, respectively.

## 3.4 Summary of Results and Possible Areas for Future Research

In this paper we prove that the distributions of likelihood ratio statistics and their signed square root can be approximated by their bootstrap distribution up to the second order $[O(1/n)]$ when the underlying sampling distribution is partly discrete. One application of this result can be applied to find accurate procedures for constructing one-sided confidence bounds, two-sided confidence intervals or joint confidence region for complete or censored data.

Examples like the one-parameter exponential model with Type I censoring and logistic regression given by Jensen (1989, 1993) illustrate some applications. We study in detail, the two-parameter Weibull distribution model when data are Type I censored. Our simulation study compares several commonly suggested methods (Bootstrap-$t$ and $BC_a$) and more accurate higher order methods (modified signed square root likelihood ratio statistic) with likelihood ratio statistics calibrated by bootstrap procedures. The simulation provides a clear view of the small sample properties of these statistics.

We can draw the following conclusions from our simulations involving Type I censored data. If one-sided confidence bounds are of interest, the PBSRLLR and PBMSRLLR methods provide better coverage probability when the expected number of failures exceeds 10. If two-sided confidence intervals are of interest, the PBSRLLR, PBMSRLLR and LLRB methods provide accurate procedures and moreover, the PBMSRLLR and LLRB methods give accurate coverage probability when the expected number of failures exceeds 7. Although the LLRB method for two-sided confidence interval is the most accurate one in coverage probability among these methods, the resulting two-sided confidence interval is not symmetric in the sense that the confidence level of one side of the interval is larger than the nominal confidence level and the confidence level of the other side of the interval is smaller than the nominal one.

Some possible areas for further research are:

- Our examples show that the theorems in Section 3.2 can be applied to the location-scale model with Type I censoring data. For other kinds of censoring and distributions, Conditions (A) and (B) can be expected to hold when the model distributions are smooth and without overly heavy tails. It would be of interest to study the finite sample coverage probabilities for such distributions.

- The procedure based on the modified likelihood ratio statistic (e.g., MSRLLR, defined in the Appendix B) provides better coverage probabilities when they are calibrated with a bootstrap procedure. The order of accuracy of the approximation could be further explored using methods parallel to those in Section 3.2.

- Although the order of accuracy is the same for different parameters of interest in the theorem, our simulation study shows that, in small samples, the accuracy of the bootstrap methods for constructing one-sided confidence bounds are quite different for different quantiles when Type I censored data are considered. The problem is in the place when the quantity of interest is the $p$ quantile where $p$ is close to the proportion failing. The reason for the problem is due to the discrete-like behavior of MLE in Type I censored data (see Jeng and Meeker (1998) for more discussion and examples on this point). When the expected number of failures is small (less than 10), another alternative suggested by some limited simulation results is to use a double bootstrap calibration. Both the theoretical and the finite sample properties of this approach could be studied. The computational effort needed to do a complete simulation experiment would, however, be extremely large.

## References

Babu, G. J. (1991). "Edgeworth expansions for statistics which are functions of lattice and non-lattice variables." *Statistics & Probability Letters*, **12**, 1-7.

Barndorff-Nielsen, O. E. (1986). "Inference on full or partial parameters based on the standardized signed log likelihood ratio." *Biometrika*, **73**, 307-322.

Barndorff-Nielsen, O. E. (1991). "Modified signed log likelihood ratio." *Biometrika*, **78**, 557-563.

Barndorff-Nielsen, O. E., and Cox, D. R. (1994). *Inference and Asymptotics*. London: Chapman & Hall

Bartlett, M. S. (1937). "Properties of sufficiency and statistical tests." *Proceedings of Royal Society of London* A **160**, 268-282.

Box, G. E. P. (1949). "A general distribution theory for a class of likelihood criteria." *Biometrika*, **36**, 317-346.

Beyer, W. H. (1979). *Standard Mathematical Tables*, 25th Edition. CRC Press Inc.

Chandra, T. K. and Ghosh, J. K. (1979). "Valid asymptotic expansions for the likelihood ratio statistics and other perturbed chi-square variables." *Sankhyā* A, **41**, 22-47.

Datta, S. (1992). "A note on the Edgeworth expansions and the bootstrap." *Sankhyā A*, **54**, 171-182.

Doganaksoy, N., and Schmee, J. (1993). "Comparisons of approximate confidence intervals for distributions used in life-data analysis." *Technometrics*, **35**, 175-184.

Efron, B. (1985). "Bootstrap confidence intervals for a class of parameter problems." *Biometrika*, **72**, 45-58.

Efron, B., and Tibshirani, R. J. (1993). *An Introduction to the Bootstrap*. New York: Chapman & Hall.

Hall, P. (1992). *The Bootstrap and Edgeworth Expansion*. New York: Springer-Verlag.

Hayakawa, T. (1977). "The likelihood ratio criterion and the asymptotic expansion of its distribution." *Annals of the Institute of Statistical Mathematics*, **29**, 359-378.

Jeng, S. L. and Meeker, W. Q. (1998). "Bootstrap likelihood ratio statistics". Preprint, Department of Statistics, Iowa State University.

Jensen, J. L. (1987). "Standardized log-likelihood ratio statistics for mixtures of discrete and continuous observations." *The Annals of Statistics*, **15**, 314-324.

Jensen, J. L. (1989). "Validity of the formal Edgeworth expansion when the underlying distribution is partly discrete." *Probability Theory and Related Fields*, **81**, 507-519.

Jensen, J. L. (1993). "A historical sketch and some new results on the improved log likelihood ratio statistic." *Scandinavian Journal of Statistics, Theory and Applications*, **20**, 1-15.

Lawley, D. N. (1956). "A general method for approximating to the distribution of the likelihood ratio criteria." *Biometrika*, **43**, 295-303.

McCullagh, P. (1984) " Local sufficiency." *Biometrika*, **71**, 233-244.

Nishii, R. and Yanagimoto, T. (1993). " Normal approximation to the distribution of the sample mean in the exponential family". *Statistical sciences and data analysis*, (Tokyo, 1991), 313-324, VSP, Utrecht.

Sargan, J. D. (1976). "Econometric estimators and the Edgeworth approximation." *Econometrica*, **44**, 421-448.

Shao, J., and Tu, D. (1995). *The Jackknife and Bootstrap*. New York: Springer-Verlag.

# 4 SIMULTANEOUS PARAMETRIC CONFIDENCE BANDS FOR CUMULATIVE DISTRIBUTIONS FROM LIFE DATA

A paper to be submitted to the Journal of Life Data Analysis

Shuen-Lin Jeng and William Q. Meeker

## Abstract

This paper describes existing methods and develops new methods for constructing simultaneous confidence bands for a cumulative distribution function (cdf). Our results are built on extensions of previous work by Cheng and Iles (1983, 1988). Cheng and Iles use Wald statistics with (expected) Fisher information and provide different approaches to find one-sided and two-sided simultaneous confidence bands. We consider three statistics, Wald statistics with Fisher information, Wald statistics with local information, and likelihood ratio statistics. Unlike pointwise confidence intervals, it is not possible to combine two 95% one-sided simultaneous confidence bands to get a 90% two-sided simultaneous confidence band. We present a general approach for construction of two-sided simultaneous confidence bands on a cdf for a continuous parametric model from complete and censored data. We start by using standard large-sample approximations and then extend and compare these to corresponding simulation or bootstrap calibrated versions of the same methods. We show that bootstrap methods provide more accurate coverage probabilities than those based on the usual large sample approximations. Both two-sided and one-sided simultaneous confidence bands for location-scale parameter model are discussed in detail including situations with complete and censored data. A simulation for the Weibull distribution and Type I censored data is used to compare finite sample properties. For the location-scale model with complete or Type II censoring, the bootstrap methods are exact. Simulation results show that, with Type I censoring, a bootstrap method based on the Wald statistic with local information provides a confidence region with coverage probability that is more accurate than a method

based on bootstraping the likelihood ratio statistic. We illustrate the implementation of the methods with an application to estimate probability of detection (POD) which is used to asses nondestructive evaluation (NDE) capability.

**Keywords**: Bootstrap, likelihood ratio, simultaneous confidence band, life data, probability of detection, Wald.

## 4.1 Introduction

### 4.1.1 Problem

In life testing and reliability studies, the primary problem of interest is often to estimate an unknown cumulative distribution function (cdf). For example, sample units might be put on a life test for the purpose of estimating the proportion failing before some specific time point. Another example is the need to quantify nondestructive evaluation (NDE) capability. NDE methods are used, for example, to detect a subsurface flaws before processing expensive materials. Inputs for a risk analysis include detection capability for a range of different flaw sizes. These problems can be formulated as one where an unknown cdf is to be estimated. We will, however, use the more familiar failure time language in our general discussion.

Confidence intervals quantify the uncertainty of estimation. For example, pointwise confidence intervals with a specific confidence level can be computed for the cdf at particular times. When the interest is on the cdf over a range of time values, the procedure using the combination of these pointwise confidence intervals will not provide a simultaneous confidence band with same coverage probability. For a given confidence level, a simultaneous confidence band would be wider than the joint set of pointwise confidence intervals. This is because we use the same information from the data to do the inference for specific point of interest as we have for inference on an infinite number of points.

Unlike pointwise confidence intervals, one cannot combine two $100(1-\alpha/2)\%$ one-sided simultaneous confidence bands to get a $100(1-\alpha)\%$ two-sided simultaneous confidence band. Different procedures are needed for one-sided and two-sided cases.

Censoring often arises in life data collection. Some theoretical results for complete data do not hold for censored data. Especially for Type I censoring, the Wald and the likelihood ratio statistics no longer have the pivotal property (a pivotal statistic has a distribution that does not depend on unknown parameters) in location-scale models. Bootstrap methods, however, provide a more accurate approximate distribution when the exact distributional form is not available. Jeng and Meeker (1998) show that the bootstrap likelihood ratio procedures are generally second order accurate for complete and censored data. Simulation results in this paper show that the procedure based on the bootstrap Wald statistics with local information provide a confidence region with confidence level that appears to be as accurate as or more accurate than the procedure based on the bootstrap likelihood ratio statistics, even when the expected number of failures is small.

### 4.1.2 Literature Review

Nonparametric methods for constructing confidence bands for cdfs can, for example, be based on statistics like the Kolmogorov-Smirnov statistics. See Lehmann (1986, pp. 355-357) for definition and references to the literature. As described in Cheng and Iles (1983), however, these methods give rise to a constant (vertical) width and part of such a band will have ordinate values that are greater than one, while other parts will have ordinate values that are negative. Even if the general approach is used in a parametric setting, it makes the band unnecessarily broad in the tails. Kanofsky and Srinivasan (1972) overcome the problem under normal, exponential and uniform models by using the maximum absolute difference between the true function and an estimator of it (similar to the Kolmogorov-Smirnov statistics) and by adjusting the resulting band to obtain the required confidence level. Using the Wald statistics with (expected) Fisher information, Cheng and Iles (1983) provide an alternative general procedure that can be applied to construct simultaneous bands for any continuous function $g(\cdot;\theta)$ of the parameters $\theta$. First, a joint confidence region is constructed for the unknown parameters. Then a simultaneous confidence band is obtained by seeing how the continuous distribution $g(\cdot;\theta)$ changes as the parameters are varied within the joint confidence region. The band is two-sided and has ordinate values that lie within the range of $g$. Cheng and Iles (1988) extend the result to one-sided simultaneous confidence bands for a cdf under the location-scale model with complete data. The simultaneous confidence bands constructed in this way may be exact or conservative. We show that, for the location-scale model with complete or Type II censoring, the bootstrap methods we present for constructing the two-sided and one-sided simultaneous confidence bands are exact.

Le Cam (1990) replaced Wald's confidence ellipsoids by confidence sets based on Hellinger distance. Escobar and Meeker (1998) develop methods parallel to Cheng and Iles (1983) based on the Wald statistics with local information for two-sided bands of quantiles and cdfs and show a certain duality property.

A likelihood ratio test can be used to construct a joint confidence region (or an approximate joint confidence region) for model parameters. Generally the distribution of the likelihood ratio statistic follows a $\chi^2$ distribution to the order $O(1/n)$ for both complete and censored data (Jensen, 1993). This confidence region can produce simultaneous confidence bands for cdf's or any continuous function $g(\cdot;\theta)$. The likelihood ratio statistics are transformation invariant, unlike the Wald statistics.

In the location-scale model for complete and Type II censored data, the Wald statistic is a pivotal statistic. One can find the distribution of the Wald statistic by using simulation (or parametric bootstrap) methods. For time-censored data, the distribution of the Wald and the likelihood ratio statistics depends on the unknown proportion in the population that would fail before the fixed censoring time. The bootstrap procedure still provides a second order accurate approximation for the distribution of likelihood ratio statistics (see Jeng and Meeker, 1998).

There are some other bootstrap methods for constructing joint confidence regions that are not included in this research. Beran (1988) suggests a method called bootstrap prepivoting to find the simultaneous confidence bands for a family of parametric functions. The advantage of Beran's method is that the resulting confidence intervals are asymptotically balanced. A simultaneous confidence band of a function $g(\cdot,\theta)$ is balanced if the pointwise confidence level for the confidence statement concerning $g(x,\theta)$ remains unchanged as $x$ varies. But the prepivoting procedure usually needs a double bootstrap to make the root closer to a pivot. Hall (1992, Section 4.2) suggests a likelihood based region that requires high dimensional density estimation. Yeh and Singh (1997) propose a bootstrap balanced confidence region based on the Tukey depth. The difficulty of using this method is the large amount of computer time required to find the Tukey depth of every single point.

Meeker et al. (1995, 1996, 1997) develop a methodology to estimate Nondestructive Evaluation (NDE) capability. The methodology is based on a physical/statistical prediction model and can be used to predict probability of detection (POD) curves and other characteristics of a flaw detecting system. Sarkar et al. (1998) apply a similar method to quantify nondestructive testing inspection capability, using limited data available from destructive testing of cracks in heat exchanger tubes. Their data were right censored because of measurement saturation for large signals. They estimate a POD curve for a particular flaw detection system and provide pointwise confidence intervals for the POD curve based on the delta method and a normal approximation. We extend the results to provide simultaneous confidence bands for the POD curve. We use a bootstrap procedure to built a joint confidence region for the unknown parameters. This bootstrap procedure is similar to the one used by Robinson (1983) to construct confidence intervals for one-dimensional parameters from progressively censored data. Then the joint confidence region is used to construct a simultaneous confidence band for the POD curve.

### 4.1.3 Overview

Section 4.2 provides a general approach for constructing two-sided simultaneous confidence bands for a function $g(\cdot;\theta)$. Section 4.3 focuses on the location-scale distribution model and Section 4.4 presents the results of a simulation study using the Weibull distribution with complete and Type I censored data. Section 4.5 presents an application in which the simultaneous confidence bands are used to quantify the uncertainty in the probability of detection curve. Section 4.6 gives discussion and possible areas for future research.

## 4.2 Methods

Let $X$ be a continuous random variable with values in a set $\mathcal{D}$ (e.g., the positive real line) and let $g(x;\theta)$ be a continuous function defined on the set $\mathcal{D}$ and the $k$ dimensional parameter space of $\theta$. A random sample $x_1, \ldots, x_n$ of size $n$ is to be used to calculate a simultaneous confidence band for $g(x;\theta)$ over some specified (possibly infinite) range of $x$ values. We present a general approach for constructing two-sided simultaneous confidence bands. The method can be used for both complete and censored data. The approach extends previous results from Cheng and Iles (1983).

Now we define some notation used in this paper. Suppose first that $\mathcal{R}$ is a $100(1-\alpha)\%$ joint confidence region for the unknown parameter vector $\theta$. $\mathcal{R}$ could be obtained for the purpose of constructing either one-sided or two-sided simultaneous confidence bands. For a given function $g$, let us consider the function $y = g(x;\theta)$ in the $(x, y)$ plane for $x \in \mathcal{D}$. When $\theta$ is changing in $\mathcal{R}$, the function $g$ will cover a region, $\mathcal{B}$, on the $(x, y)$ plane. Because the true value of $\theta$ lies in $\mathcal{R}$ with probability $1-\alpha$, the probability is at least $1-\alpha$ that one of the functions used to cover the region $\mathcal{B}$ is the unknown true function $g(\cdot;\theta)$. Thus $\mathcal{B}$ is a simultaneous confidence band for $g(\cdot;\theta)$ that will contain the true function $g(\cdot;\theta)$ with probability at least $1-\alpha$. In general there may be values of $\theta$ outside of the region $\mathcal{R}$ that give a function $g(\cdot;\theta)$ lying entirely within the band $\mathcal{B}$. So the band $\mathcal{B}$ could be conservative.

Define the lower and upper confidence curves $C_l$ and $C_u$ at $x$ corresponding to a joint confidence region $\mathcal{R}$ as

$$C_l(x) = \min_{\theta \in \mathcal{R}} g(x;\theta), \quad C_u(x) = \max_{\theta \in \mathcal{R}} g(x;\theta). \tag{4.1}$$

If $\mathcal{R}$ is the region constructed for a two-sided simultaneous confidence band, we denote the two-sided band by

$$\mathcal{B} = \{(x, y) : C_l(x) \leq y \leq C_u(x), x \in \mathcal{D}\}. \tag{4.2}$$

Usually in order to achieve the required confidence level, a different joint confidence region $\mathcal{R}$ is needed for a one-sided simultaneous band. For a region $\mathcal{R}$ constructed to compute a one-sided band, we use the region to produce a lower confidence curve $C_l(x)$ and denote the one-sided lower simultaneous confidence band by

$$\mathcal{B}_l = \{(x, y) : y \geq C_l(x), x \in \mathcal{D}\}. \tag{4.3}$$

Similarly, we denote the the one-sided upper simultaneous confidence band by

$$\mathcal{B}_u = \{(x, y) : y \leq C_u(x), x \in \mathcal{D}\}. \tag{4.4}$$

### 4.2.1 Methods Used

The different methods for constructing a joint confidence region $\mathcal{R}$ are based on different statistics and procedures. Below we describe briefly seven methods by indicating how the exact or approximate distribution of the statistics are obtained. In all of these methods, we let $L(\theta)$ denote the likelihood function and $\widehat{\theta}$ denote the maximum likelihood estimator of $\theta$.

#### 4.2.1.1 $\chi^2$-approximation Methods

**Wald statistic with Fisher information (WLADF).** Let

$$I(\theta) = \mathrm{E}\left[\frac{\partial \log L(\theta)}{\partial \theta_i}\frac{\partial \log L(\theta)}{\partial \theta_j}\right]$$

be the Fisher information matrix. The Wald statistic with Fisher information is

$$Q_F(\theta) = (\widehat{\theta} - \theta)' I(\theta)(\widehat{\theta} - \theta).$$

Rao (1973, page 418) shows that the large-sample limiting distribution of $Q_F$ is $\chi^2_k$.

**Wald statistic with local information (WLADL).** Let

$$\widehat{I}(\widehat{\theta}) = \left[\frac{\partial \log L(\widehat{\theta})}{\partial \theta_i}\frac{\partial \log L(\widehat{\theta})}{\partial \theta_j}\right]$$

be the local information matrix. The Wald statistic with local information is

$$Q_L(\theta) = (\widehat{\theta} - \theta)'\widehat{I}(\widehat{\theta})(\widehat{\theta} - \theta).$$

Cox and Hinkley (1974, page 314) show that the large-sample limiting distribution of $Q_L$ is $\chi^2_k$.

**Log LR method (LLR).** The likelihood ratio statistic is defined as

$$W(\theta) = -2\log\left[\frac{L(\theta)}{L(\widehat{\theta})}\right].$$

Serfling (1980, Section 4.4) shows that the large-sample limiting distribution of $W(\theta)$ is $\chi^2_k$.

**Log LR Bartlett corrected method (LLRB).** Let

$$W_B(\theta) = k\frac{W(\theta)}{\mathrm{E}[W(\theta)]}.$$

Because the expectation of $W_B(\theta)$ is equal to the mean of the $\chi^2_k$ distribution, the distribution of $W_B(\theta)$ can, when compared with $W(\theta)$, be expected to be better approximated by the $\chi^2_k$ distribution (Bartlett 1937). In general one must substitute an estimate for $\mathrm{E}[W(\theta)]$ computed from one's data. For complicated problems (e.g., those involving censoring) it is necessary to estimate of $\mathrm{E}[W(\theta)]$ by using simulation.

#### 4.2.1.2 Parametric Bootstrap Methods

The following methods use the "bootstrap principle" or Monte Carlo evaluation of sampling distributions. Suppose a statistic $S(\theta)$ is a function of random variables with a distribution that depends on the parameter $\theta$. The parametric bootstrap version $S^*(\widehat{\theta})$ of $S$ is the same function but evaluated at data ("bootstrap samples") simulated using an estimate $\widehat{\theta}$ instead of the unknown $\theta$ [see Efron and Tibshirani (1993) for more detail]. Using $\widehat{\theta}$ in place of the distribution parameters, the distribution of $S^*$ can be calculated analytically in simple situations, or by simulation in general. Except for special cases in which the underling statistic is pivotal [e.g., complete data or Type II censoring from location-scale distributions] the distribution of $S^*$ will depend on $\hat{\theta}$, and thus the distribution of $S^*$ will provide only an approximation to the distribution of $S$.

**Parametric bootstrap Wald statistic with Fisher information (BWALDF).** Let $Q_F^*(\theta)$ be the bootstrap version of $Q_F(\theta)$. Use the distribution of $Q_F^*(\theta)$ to approximate the distribution of $Q_F(\theta)$.

**Parametric bootstrap Wald statistic with local information (BWALDL).** Let $Q_L^*(\theta)$ be the bootstrap version of $Q_L(\theta)$. Use the distribution of $Q_L^*(\theta)$ to approximate the distribution of $Q_L(\theta)$.

**Parametric bootstrap log likelihood ratio method (BLLR).** Let $W^*(\theta)$ be the bootstrap version of $W(\theta)$. Use the distribution of $W^*(\theta)$ to approximate the distribution of $W(\theta)$.

### 4.2.2 Construction of Simultaneous Confidence Bands

Let $S(\theta)$ be any one of $Q_L(\theta)$, $Q_F(\theta)$, $W(\theta)$, or $W_B(\theta)$. Also let $\gamma$ denote the $100(1-\alpha)\%$ quantile from the distributions corresponding to $S(\theta)$ or $S^*(\theta)$ from one of the seven methods. A $100(1-\alpha)\%$ confidence region for $\theta$ can be obtained by

$$\mathcal{R} = \{\theta : S(\theta) < \gamma\}. \tag{4.5}$$

By using the notation of (4.1), an approximate $100(1-\alpha)\%$ two-sided simultaneous confidence band can be obtained from (4.2).

In general, these methods provide exact or conservative procedures. In the next section we show that under the location-scale model with complete or Type II censored data, the coverage probability of the bootstrap methods for constructing two-sided simultaneous confidence bands for the cdf by using the joint confidence region $\mathcal{R}$ of $\theta$ is equal to the nominal confidence level. When the data are Type I censored from a location-scale model or when the data are complete from other general models, approximate two-sided simultaneous confidence bands still can be constructed using these methods.

The confidence level for a one-sided confidence band constructed by using the joint confidence region obtained with equation (4.5) will be larger than the nominal one. In general the joint confidence region needed for the one-sided simultaneous confidence bands depends on the properties of function $g$. However, for the location-scale model with complete data, Cheng and Iles (1988) give a procedure that provides a one-sided bound with the correct coverage probability.

Jeng and Meeker (1998) show that under some regularities conditions the BLLR method is second order accurate even for Type I censored data. This procedure, when used to construct joint confidence regions, has coverage probability that is close to nominal. The BLLR method also provides approximate two-sided and one-sided simultaneous confidence bands for general models with complete or censored data.

In the next section we focus on methods for the location-scale model with complete or censored data. These methods can, however, be applied to log-location-scale distributions like the lognormal, Weibull and loglogistic.

## 4.3 Location-scale Model

Suppose $\Phi(\xi)$ is a known continuous distribution function, and consider a random variable $X$ with cdf $\Phi[(x-\mu)/\sigma]$ and density $\phi[(x-\mu)/\sigma]/\sigma$ where $\mu$ and $\sigma$ are the unknown location and scale parameters. In this case $X$ is said to have a location-scale distribution. Let $\widehat{\mu}$ and $\widehat{\sigma}$ be the maximum likelihood estimators for $\mu$ and $\sigma$.

### 4.3.1 Two-sided Simultaneous Confidence Bands

This section describes some properties of the statistics that are used to construct joint confidence regions. First we express these statistics into different forms.

**WALDF**. The Fisher information matrix for $\mu$ and $\sigma$ can be written as

$$I(\mu,\sigma) = \frac{n}{\sigma^2}\begin{pmatrix} i_{11} & -i_{12} \\ -i_{12} & i_{22} \end{pmatrix}.$$

Then, as shown in Cheng and Iles (1988), the Wald statistic with Fisher information can be expressed as

$$\begin{aligned} Q_F(\mu,\sigma) &= ni_{11}(\widehat{\mu}-\mu)^2/\sigma^2 - 2ni_{12}(\widehat{\mu}-\mu)(\widehat{\sigma}-\sigma)/\sigma^2 + ni_{22}(\widehat{\sigma}-\sigma)^2/\sigma^2 \\ &= ni_{11}M^2 - 2ni_{12}MS + ni_{22}S^2, \end{aligned}$$

where $M = (\widehat{\mu}-\mu)/\sigma$ and $S = (\widehat{\sigma}-\sigma)/\sigma$.

**WALDL**. Similarly the local information matrix can be written as

$$\widehat{I} = I(\widehat{\mu},\widehat{\sigma}) = \frac{n}{\widehat{\sigma}^2}\begin{pmatrix} \widehat{i}_{11} & -\widehat{i}_{12} \\ -\widehat{i}_{12} & \widehat{i}_{22}. \end{pmatrix}.$$

Then the Wald statistic with local information can be expressed as

$$Q_L(\mu,\sigma) = n\widehat{i}_{11}\widehat{M}^2 - 2n\widehat{i}_{12}\widehat{M}\widehat{S} + n\widehat{i}_{22}\widehat{S}^2.$$

where $\widehat{M} = (\widehat{\mu} - \mu)/\widehat{\sigma}$ and $\widehat{S} = (\widehat{\sigma} - \sigma)/\widehat{\sigma}$.

**LLR.** For complete data, the likelihood ratio statistic can be expressed as

$$W(\mu, \sigma) = -2\log\left[\left(\frac{\widehat{\sigma}}{\sigma}\right)^n \frac{\prod_{i=1}^n \phi\left(\frac{x_i - \mu}{\sigma}\right)}{\prod_{i=1}^n \phi\left(\frac{x_i - \widehat{\mu}}{\widehat{\sigma}}\right)}\right]$$

For right censored data (Type I or Type II), let $\delta_i = 1$ if the $i$th observation is a failure, $\delta_i = 0$ if $i$th the observation is censored. Then the likelihood ratio statistics can be expressed as

$$W(\mu, \sigma) = -2\log\left\{\left(\frac{\widehat{\sigma}}{\sigma}\right)^{\sum_{i=1}^n \delta_i} \frac{\prod_{i=1}^n \phi\left(\frac{x_i - \mu}{\sigma}\right)^{\delta_i}\left[1 - \Phi\left(\frac{x_i - \mu}{\sigma}\right)\right]^{1-\delta_i}}{\prod_{i=1}^n \phi\left(\frac{x_i - \widehat{\mu}}{\widehat{\sigma}}\right)^{\delta_i}\left[1 - \Phi\left(\frac{x_i - \widehat{\mu}}{\widehat{\sigma}}\right)\right]^{1-\delta_i}}\right\}.$$

**Complete data.** Kendall and Stuart (1979) show that $i_{11}, i_{12}$, and $i_{22}$ are constants independent of $\mu$ and $\sigma$. Because the distributions of $M$ and $S$ do not depend on $\mu$ and $\sigma$ (Lawless 1982, page 147), $Q_F$ is a pivotal quantity. Note that $(x_i - \widehat{\mu})/\widehat{\sigma}$, $\widehat{M}$, and $\widehat{S}$ are functions of $\widehat{\sigma}/\sigma$, $(x_i - \mu)/\sigma$, $M$, and $S$, so the distribution of $(x_i - \widehat{\mu})/\widehat{\sigma}$, $\widehat{M}$, and $\widehat{S}$ do not depend on $\mu$ or $\sigma$. The elements $\widehat{i}_{11}, \widehat{i}_{12}$, and $\widehat{i}_{22}$ depend only on $(x_i - \widehat{\mu})/\widehat{\sigma}$. Thus $Q_L$ is also a pivotal quantity. Because $W$ depends only on $(x_i - \mu)/\sigma$, $(x_i - \widehat{\mu})/\widehat{\sigma}$, and $\widehat{\sigma}/\sigma$, it is also a pivotal quantity. The BWALDL, BWALDF, and BLLR methods for constructing the $100(1-\alpha)\%$ confidence regions have exact confidence level $1-\alpha$ except for the Monte Carlo simulation error (which can be made arbitrary small by increasing the number of Monte Carlo trails).

**Type II censored data.** Lawless (1982, page 147) shows that, with Type II censoring, $Z_1 = (\mu - \widehat{\mu})/\widehat{\sigma}$, $Z_2 = \widehat{\sigma}/\sigma$, $Z_3 = (\mu - \widehat{\mu})/\sigma$, and $a_i = (x_i - \widehat{\mu})/\sigma$ are pivotal quantities. Because $Q_F$, $Q_L$, and $W$ only depend on $Z_1$, $Z_2$, $Z_3$, and $a_i$, they are also pivotal quantities with Type II censoring. The $100(1-\alpha)\%$ confidence regions obtained by the BWALDL, BWALDF, and BLLR methods have exact coverage probability $1-\alpha$ except for Monte Carlo simulation error.

**Type I censored data.** With Type I censoring the distributions of $Q_F$, $Q_L$, $W$, and $W_B$ depend on the unknown proportion failing at the censoring time. For this reason, joint confidence regions and simultaneous confidence bands based on these statistics are only approximations. The approximation improves with increasing sample size.

Once we have a $100(1-\alpha)\%$ confidence region $\mathcal{R}$ for $\mu$ and $\sigma$ from one of the previous described methods, the $100(1-\alpha)\%$ two-sided simultaneous confidence curves can be obtained by using the equation (4.1), providing the simultaneous band as indicated in (4.2).

Cheng and Iles (1983) show that the confidence level of the two-sided simultaneous confidence band is the same for the location-scale model as the confidence level of the confidence region produced by the WALDF method. We extend this result to show that any convex confidence region with required confidence level can be used to construct a two-sided simultaneous band and that both the region and the band have the required confidence level. Note that the confidence regions for the two-sided simultaneous bands constructed from the WALDF or the WALDL methods are ellipses and thus are convex. Usually the LLR method will produce convex confidence regions.

We first use a convex confidence region $\mathcal{R}$ to construct a simultaneous confidence band for quantiles of the distribution. Then we show that the band can be converted to a simultaneous confidence band for the cdf and argue that in either case, the confidence level of the band is the same as the confidence region $\mathcal{R}$.

The $p$ quantile $x_p$ is defined as

$$x_p = \mu + u_p\sigma. \tag{4.6}$$

where $u_p = \Phi^{-1}(p)$. Consider a fixed $p$, $0 < p < 1$. In the $(\mu, \sigma)$ plane, equation (4.6) represents a family of parallel lines with different intercepts $x_p$ and the same slope $-u_p^{-1}$. Because the region $\mathcal{R}$ is convex, the smallest and the largest values of $x_p$ produced by $(\mu, \sigma) \in \mathcal{R}$, say $\widehat{x}_p(\min)$ and $\widehat{x}_p(\max)$, correspond to two parallel tangents to the region $\mathcal{R}$ (see Figure 4.1). Then $[\widehat{x}_p(\min), p]$ and $[\widehat{x}_p(\max), p]$, $0 < p < 1$, are two curves in the $(x, \Phi[(x-\mu)/\sigma])$ plane which define a simultaneous confidence band $\mathcal{B}$ for all quantiles.

Based on Result 1 in Appendix C, the lower and upper confidence curves for quantiles are the same as the upper and lower confidence curves, respectively, for the cdf $\Phi$. That is the band $\mathcal{B}$ is also the simultaneous confidence band for the cdf. Result 2 in Appendix C shows the band $\mathcal{B}$ has the same confidence level as the confidence region $\mathcal{R}$.

Equation (4.1) can always be calculated numerically, so the band (4.2) is obtainable. Cheng and Iles (1983) provide an exact formula for the WALDF method. Suppose $\gamma$ is the $1-\alpha$ quantile of the distribution from the WALDF method and $\gamma/n < (i_{11}i_{22} - i_{12}^2)/i_{11}$. Then $100(1-\alpha)\%$ two-sided simultaneous confidence curves are

$$C_l(x) = \Phi(\widehat{\xi}(x) - h), \quad C_u(x) = \Phi(\widehat{\xi}(x) + h) \tag{4.7}$$

where

$$h = \{\gamma n^{-1} i_{11}^{-1}[1 + (i_{11}\widehat{\xi} + i_{12})^2(i_{11}i_{22} - i_{12}^2)^{-1}]\}^{1/2} \text{ and } \widehat{\xi}(x) = (x - \widehat{\mu})/\widehat{\sigma}.$$

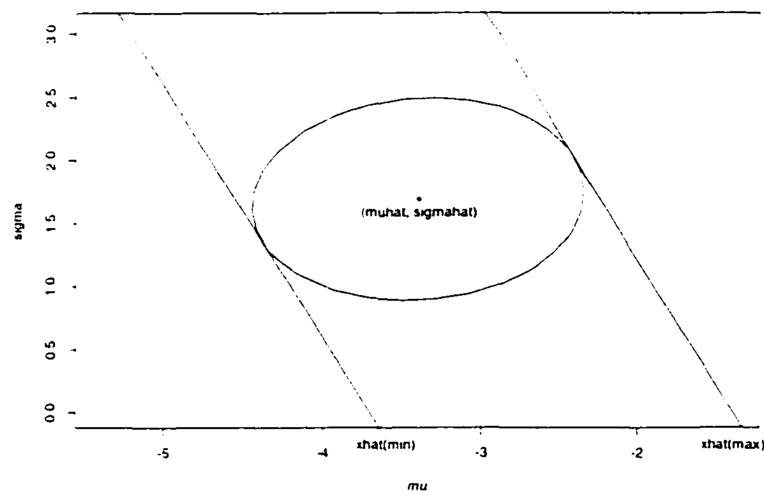



Figure 4.1 A 95% convex confidence region to be used for two-sided simultaneous confidence band constructed from the BWALDL method with data in Section 4.5.

The exact same formula can be applied for the BWALDF method, except that the value of $\gamma$ is replaced by $\gamma^*$, the $1-\alpha$ quantile of distribution of $Q_L^*$.

Using the arguments similar to those in Cheng and Iles (1983), Escobar and Meeker (1998) develop the following formula for the WALDL method. Their formula also can be used for the BWALDL method. Suppose $\gamma$ is the $1-\alpha$ quantile of the distribution from the WALDL or BWALDL method and $\gamma/\widehat{\sigma}^2 < (\widehat{i}_{11}\widehat{i}_{22} - \widehat{i}_{12}^2)/\widehat{i}_{11}$. Then $100(1-\alpha)\%$ two-sided simultaneous confidence curves are

$$C_l(x) = \Phi(\widehat{\xi}(x) + h_1 - h_2), \quad C_u(x) = \Phi(\widehat{\xi}(x) + h_1 + h_2) \tag{4.8}$$

where

$$\begin{aligned} h_1 &= \frac{\gamma \times (\widehat{i}_{12} + \widehat{\xi}\,\widehat{i}_{22})}{\widehat{\sigma}^2 - \gamma \times \widehat{i}_{22}} \\ h_2 &= \frac{\sqrt{\gamma}}{\widehat{\sigma}^2 - \gamma \times \widehat{i}_{22}} \sqrt{\sigma^2 \times (\widehat{i}_{11} + 2\widehat{\xi}\,\widehat{i}_{12} + \widehat{\xi}^2\,\widehat{i}_{22}) - \gamma \times (\widehat{i}_{11}\widehat{i}_{22} - \widehat{i}_{12}^2)} \\ \widehat{\xi}(x) &= (x - \widehat{\mu})/\widehat{\sigma}. \end{aligned}$$

For the LLR, LLRB, and BLLR methods, the $100(1-\alpha)\%$ two-sided simultaneous confidence curves can be obtained numerically by using Equation (4.1).

### 4.3.2 One-sided Simultaneous Confidence Bands

The construction of a confidence region for one-sided simultaneous confidence bands is different from the two-sided case. Cheng and Iles (1988) provide an argument for using the WALDF method. We extend their argument to other methods that can be used to produce convex confidence regions.

To see this, we describe a method for obtaining a region needed to define a lower confidence band of the cdf (the method for an upper confidence band is analogous). As argued in Section 4.3.1, the upper simultaneous confidence curve for quantiles is the same as the lower simultaneous confidence curve for the cdf. The same argument applies for one-sided simultaneous confidence bands. Below we construct a confidence region for obtaining an upper simultaneous confidence band for quantiles and argue that the confidence level of the band is the same as that of the region.

Suppose we have a convex confidence region $\mathcal{R}$ with a certain confidence level. For a given $p$, let $\mathcal{R}_p$ denote the half space of $(\mu, \sigma)$ that satisfies $\mu + u_p\sigma \leq \widehat{x}_p(\max)$. Let $\mathcal{R}_l$ denote the intersection of all $\mathcal{R}_p$, $0 < p < 1$. Because the tangent lines are on the right boundary of $\mathcal{R}$, $\mathcal{R}_l$ is the union of region $\mathcal{R}$ and a left semi-infinite band $\mathcal{S}_l$. See Figure 4.2.

Result 3 (in Appendix C) shows that the confidence level of $\mathcal{B}_l$ obtained by using $C_l$ is the same as that of $\mathcal{R}_l$. That is, the one-sided simultaneous confidence band will be exact if the corresponding convex confidence region $\mathcal{R}_l$ has the desired confidence level.

We consider $\mathcal{R}_l$, $\mathcal{R}$, and $\mathcal{S}_l$ in their inverted form as being a region $\mathcal{R}_l'$, $\mathcal{R}'$, and $\mathcal{S}_l'$ in the $(\widehat{\mu}, \widehat{\sigma})$ plane. For a given confidence coefficient $1-\alpha$, we would like to calculate the corresponding value $\gamma$ such that

$$\Pr[(\widehat{\mu}, \widehat{\sigma}) \in \mathcal{R}_l' = \mathcal{R}' \cup \mathcal{S}_l'] = 1 - \alpha. \tag{4.9}$$

For the WALDF method, Cheng and Iles (1988) describe a way to calculate the critical value $\gamma$. We show their results here. Let

$$\begin{pmatrix} \widehat{\theta}_{F1} \\ \widehat{\theta}_{F2} \end{pmatrix} = \frac{1}{\sigma\sqrt{n i_{ii}}} \begin{pmatrix} i_{11} & -i_{12} \\ 0 & d \end{pmatrix} \begin{pmatrix} \widehat{\mu} - \mu \\ \widehat{\sigma} - \sigma \end{pmatrix},$$

where $d = (i_{11} i_{22} - i_{12}^2)^{1/2}$. Note that $\widehat{\theta}_{F1}$ and $\widehat{\theta}_{F2}$ are asymptotically independently normal distributed. The region $\mathcal{R}$ is defined by those $(\mu, \sigma)$ values that satisfy the inequality $Q_F = \widehat{\theta}_{F1}^2 + \widehat{\theta}_{F2}^2 \leq \gamma$. Then spherical symmetry of the independent bivariate normal distribution allows $\Pr\{(\widehat{\mu}, \widehat{\sigma}) \in \mathcal{R}_l'\}$ to be evaluated as half the sum of $\Pr\{(\widehat{\mu}, \widehat{\sigma}) \in \mathcal{R}'\}$ and $\Pr\{(\widehat{\mu}, \widehat{\sigma}) \in \mathcal{S}_d'\}$ where $\mathcal{S}_d'$ is the doubly infinite band defined by $|\widehat{\theta}_{F2}| \leq \gamma^{1/2}$.

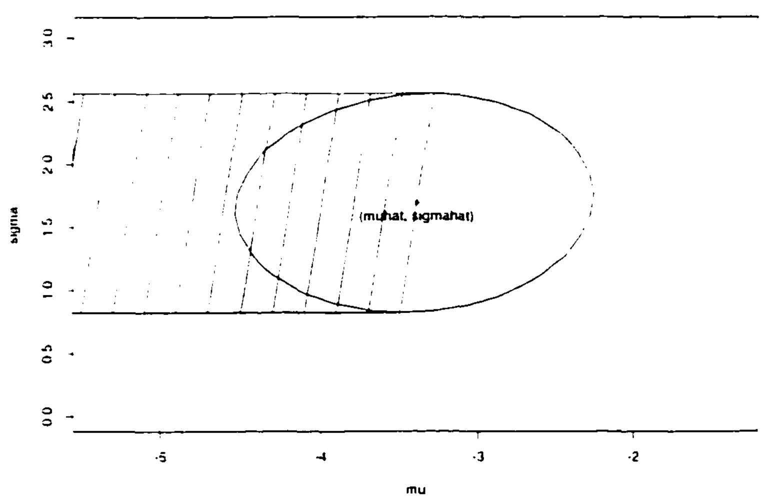



Figure 4.2 The 97.5% convex confidence region for a one-sided simultaneous confidence band constructed from the BWALDL method with data in Section 4.5. It is the union of a closed convex region and a left semi-infinite band.

These probabilities are, respectively, $\Pr(Z_1 \leq \gamma)$ and $\Pr(-\gamma^{1/2} \leq Z_2 \leq \gamma^{1/2})$, where $Z_1$ is a chi-square random variable with 2 df and $Z_2$ is a standard normal random variable. Thus the asymptotic value of the confidence coefficient associated with the region $\mathcal{R}_l$ is given by the formula $\frac{1}{2}[\Psi(\gamma) + 2\Phi_{\text{nor}}(\gamma^{1/2}) - 1]$, where $\Psi$ is the cdf of $\chi^2_2$ distribution and $\Phi_{\text{nor}}$ is the cdf of the standard distribution. For the WALDF method, to find the $\gamma$ for an approximate $100(1-\alpha)\%$ confidence region $\mathcal{R}_l$ we solve the equation

$$\frac{1}{2}[\Psi(\gamma) + 2\Phi_{\text{nor}}(\gamma^{1/2}) - 1] = 1 - \alpha. \tag{4.10}$$

For the WALDL method, let

$$\begin{pmatrix} \widehat{\theta}_{L1} \\ \widehat{\theta}_{L2} \end{pmatrix} = \frac{1}{\widehat{\sigma}\sqrt{n\widehat{i}_{11}}} \begin{pmatrix} \widehat{i}_{11} & -\widehat{i}_{12} \\ 0 & \widehat{d} \end{pmatrix} \begin{pmatrix} \widehat{\mu} - \mu \\ \widehat{\sigma} - \sigma \end{pmatrix},$$

where $\widehat{d} = (\widehat{i}_{11}\widehat{i}_{22} - \widehat{i}^2_{12})^{1/2}$. By the same argument, we solve the equation (4.10) to find $\gamma$ for an approximate $100(1-\alpha)\%$ confidence region $\mathcal{R}_l$

For the LLR method, the regions $\mathcal{R}'$ and $\mathcal{S}'_d$ are defined by $W' \leq \gamma$ and $-\gamma^{1/2} \leq R_s \leq \gamma^{1/2}$, where $R_s = \text{sign}(\widehat{\sigma} - \sigma)\sqrt{W_1}$, $W_1 = 2[\log L(\widehat{\mu}, \widehat{\sigma}) - \log L(\widehat{\mu}_\sigma, \sigma)]$, $\widehat{\mu}_\sigma$ is the constrained maximum likelihood estimator of $\mu$ given $\sigma$. Because $W$ and $R_s$ are asymptotically $\chi^2_2$ and standard normal distributed respectively, we can use equation (4.10) to find the $\gamma$ for an approximate $100(1-\alpha)\%$ confidence region $\mathcal{R}_l$. Note that because both the LLRB and the LLR statistics have the same limiting distribution, the LLRB method has the same $\gamma$ value as the LLR method.

Based on experience with pointwise confidence intervals (e.g., Jeng and Meeker, 1998), we expect that using bootstrap calibration to obtain $\gamma$ in the one-sided case will provide a more accurate procedure. Let $\widehat{\mu}^*$ and $\widehat{\sigma}^*$, $\mathcal{R}_l'^*$, $\mathcal{R}'^*$, and $\mathcal{S}_l'^*$ be the bootstrap versions of $\widehat{\mu}$ and $\widehat{\sigma}$, $\mathcal{R}_l'$, $\mathcal{R}'$, and $\mathcal{S}_l'$, respectively. Now $\mathcal{R}'$ is defined by $Q_F$, $Q_L$, $W$ or $W_B$ as (4.5). For a given confidence coefficient $1-\alpha$, we would like to calculate the corresponding value $\gamma^*$ such that

$$\Pr[(\widehat{\mu}^*, \widehat{\sigma}^*) \in \mathcal{R}_l'^* = \mathcal{R}'^* \cup \mathcal{S}_l'^*] = 1-\alpha. \tag{4.11}$$

Then we use $\gamma^*$ in place of $\gamma$ in the WALDF, WALDL, and LLR methods to provide bootstrap confidence regions.

Once the confidence region is constructed, the lower one-sided confidence curve for cdf $\Phi$ is $C_l(x) = \min_{\theta \in R_l} \Phi(x;\theta)$. Hence the the lower one-sided confidence band is given by equation (4.3).

Using arguments similar to those in the previous section, $\widehat{\theta}_{F2}$, $\widehat{\theta}_{L2}$, and $R_s$ are pivotal quantities for complete and Type II censored data. Then the confidence region obtained by bootstrap calibration has exactly the nominal confidence level (except for the Monte Carlo simulation error). Thus the procedure for one-sided simultaneous confidence bands also has the correct coverage probability. For Type I censored data, again we have only approximate results, with the approximation becoming better in large samples.

For calculation of one-sided simultaneous confidence curves, (4.7) can be used for the WALDF and the BWALDF methods by substituting in the corresponding $\gamma$ values. Formula (4.7) can also be used for the WALDL and the BWALDL methods. For the LLR and BLLR method, there is no simple formula but the one-sided simultaneous confidence curves can be calculated numerically from (4.1).

## 4.4 Simulation Study

To explore the finite sample performance of these methods, we conducted a simulation using the Weibull distribution and both complete and Type I censored data. Our simulation experiment was designed to study the following factors:

Table 4.1 Number of the cases where $r = 0$ or 1 in 5000 Monte Carlo simulations of the experiment. The expected numbers rounded to the nearest integer are shown inside parentheses.

| | | $p_f$ .01 | .10 | .50 | .90 |
|---|---|---|---|---|---|
| | 3 | 984(988) | 889(918) | 555(546) | 132(139) |
| $E(r)$ | 5 | 175(198) | 159(168) | 54( 53) | 1( 2) |
| | 7 | 34( 35) | 24( 27) | 6( 4) | 0( 0) |
| | 10 | 2( 2) | 2( 1) | 0( 0) | 0( 0) |

- $p_f$: the expected proportion failing by the censoring time.
- $E(r) = np_f$: the expected number of failures before the censoring time.

We used 5000 Monte Carlo samples for each $p_f$ and $E(r)$ combination. The number of bootstrap replications used was 10000. The levels of the simulation experiment factors used were $p_f = .01, .1, .5, .9, 1$ and $E(r) = 3, 5, 7, 10, 15, 20$, and 30. For each Monte Carlo sample we obtained the ML estimates of the location and scale parameters. The confidence regions for the two-sided and one-sided $100(1-\alpha)\%$ simultaneous confidence bands were evaluated for $\alpha$ =.025 and .05. Without loss of generality, we sampled from an SEV distribution with $\mu = 0$ and $\sigma = 1$.

Because the number of failures before the censoring time $t_c$ is random, it is possible to have as few as $r = 0$ or 1 failures in the simulation, especially when $E(r)$ is small. With $r = 0$, ML estimates do not exist. With $r = 1$, LR intervals may not exist. Therefore, we calculate the results conditionally on the cases with $r > 1$, and report the observed nonzero proportions that resulted in $r \leq 1$. See Table 4.4.

Let $1-\alpha$ be the nominal coverage probability (CP) of a procedure for constructing a joint confidence region, and let $1-\hat{\alpha}$ denote the corresponding Monte Carlo evaluation of the actual coverage probability $1-\alpha'$. The standard error of $\hat{\alpha}$ is approximately $\text{se}(1-\hat{\alpha}) = [\alpha'(1-\alpha')/n_s]^{1/2}$, where $n_s$ is the number of Monte Carlo simulation trials. For a 95% confidence region from 5000 simulations the standard error of the Monte Carlo CP evaluation is $[.05(1-.95)/5000]^{1/2} = .0031$ if the procedure is correct. Thus the Monte Carlo error is approximately $\pm 1\%$. We say the procedure or the method for the 95% confidence region is adequate if the CP is within $\pm 1\%$ error of the nominal CP.

From the Figure 4.3 to Figure 4.7 we have the following results

- Neither the WALDF nor the WALDL method provides an adequate procedure when $E(r) \leq 30$.
- The coverage probability of the LLR method depends both on the sample size and on the expected number of failures. The procedure is adequate when $E(r) \geq 20$.
- The LLRB method is adequate when $E(r) \geq 5$ for two-sided simultaneous bands and when $E(r) \geq 30$ for one-sided simultaneous bands. Using a Bartlett correction improves the coverage probability of the procedure for one-sided simultaneous bands only when there is no censoring or slight censoring.
- As expected, the BLLR method is exact for complete data. The coverage probability of the procedure for two-sided simultaneous confidence bands is accurate even in heavily censored cases when $E(r) = 5$. But the coverage probability of the procedure for one-sided simultaneous confidence bands is accurate when $E(r) \geq 15$.
- As expected, the BWALDF method is exact for complete data. With Type I censoring it is adequate when $E(r) \geq 10$.
- The BWALDL is exact as expected for complete data. With Type I censoring it is adequate when $E(r) \geq 5$.

Overall, the BWALDL method provides the best results. Also there are simple formulas to calculate simultaneous confidence bands for the cdf. So for the location-scale mode, the BWALDL method is recommended.

## 4.5 Simultaneous Confidence Band for POD Curve

Probability of Detection (POD) curves are a commonly used metric for the Nondestructive Evaluation (NDE) capability. We follow the methodology developed by Meeker et al. (1995, 1996, 1997) which is motivated by the need for methods to predict ultrasonic (UT) inspection POD for detecting hard-alpha and other subsurface flaws in titanium using a gated peak-to-peak UT detection method. Sarkar et al. (1998) apply a similar methodology to the non-destructive testing using UT inspection and destructive testing of cracks in heat exchanger data.

The combined data used in Sarkar et al. (1998) came from testing three heat exchanger tubes. The data are denoted by $\{(a_k, y_k) : k = 1, \ldots, n\}$, where $y_k$ is the signal amplitude corresponding to the crack size $a_k$. In the ultrasonic inspection, the signal

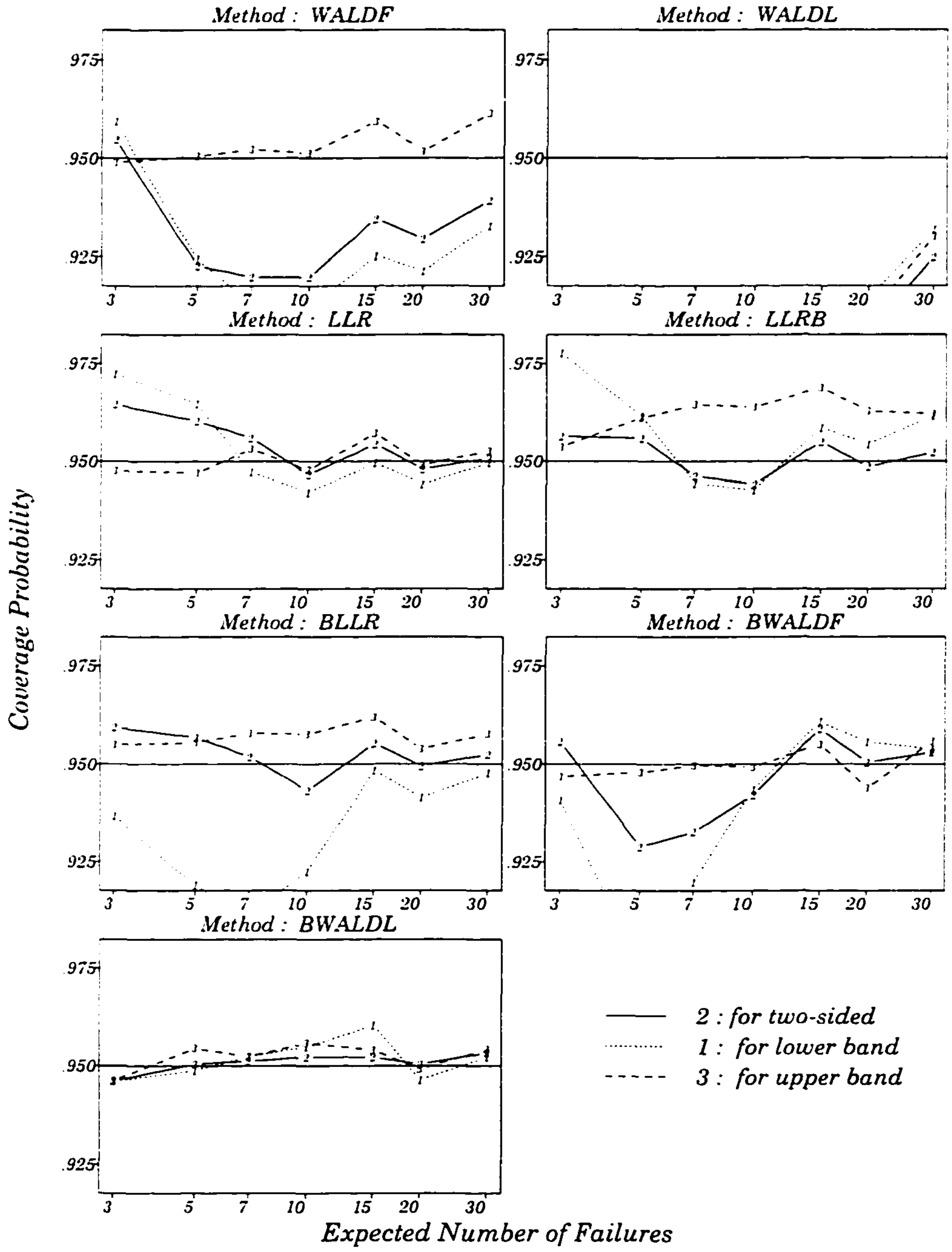



Figure 4.3 Coverage probability plot of the methods for constructing approximate 95% two-sided and one-sided simultaneous confidence bands in the cases $p_f = .01$.

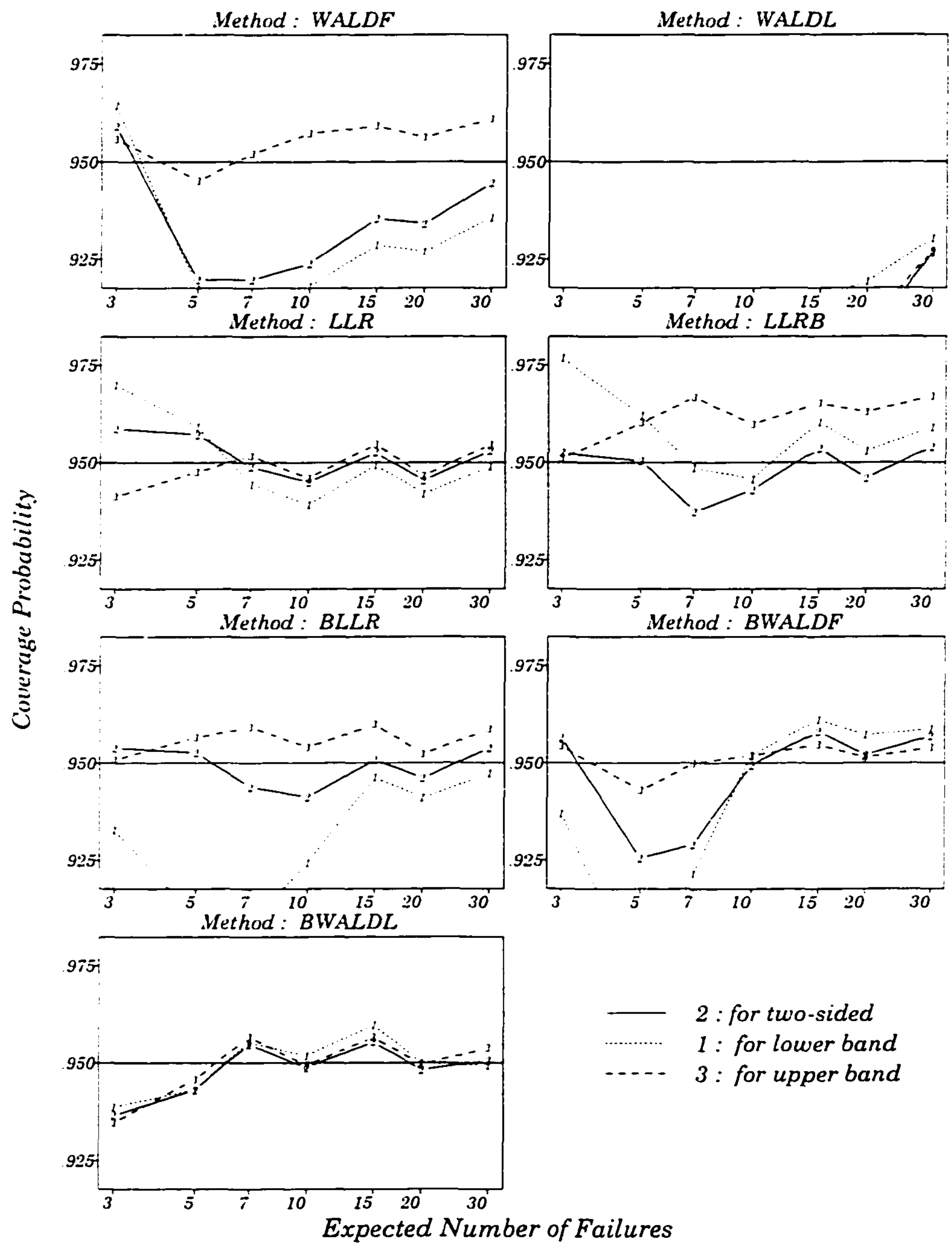



Figure 4.4 Coverage probability plot of the methods for constructing approximate 95% two-sided and one-sided simultaneous confidence bands in the cases $p_f = .1$.

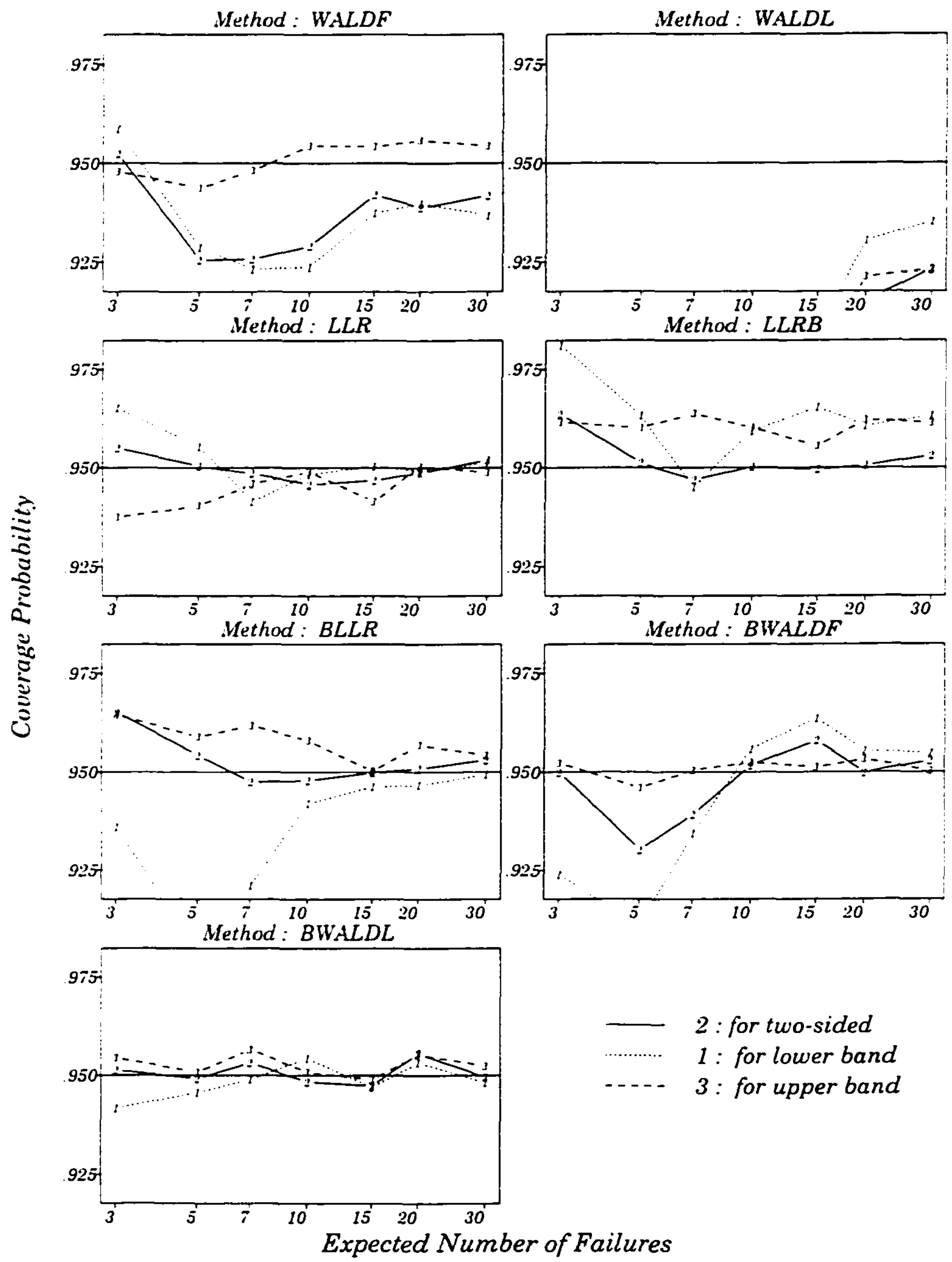



Figure 4.5 Coverage probability plot of the methods for constructing approximate 95% two-sided and one-sided simultaneous confidence bands in the cases $p_f = .5$.

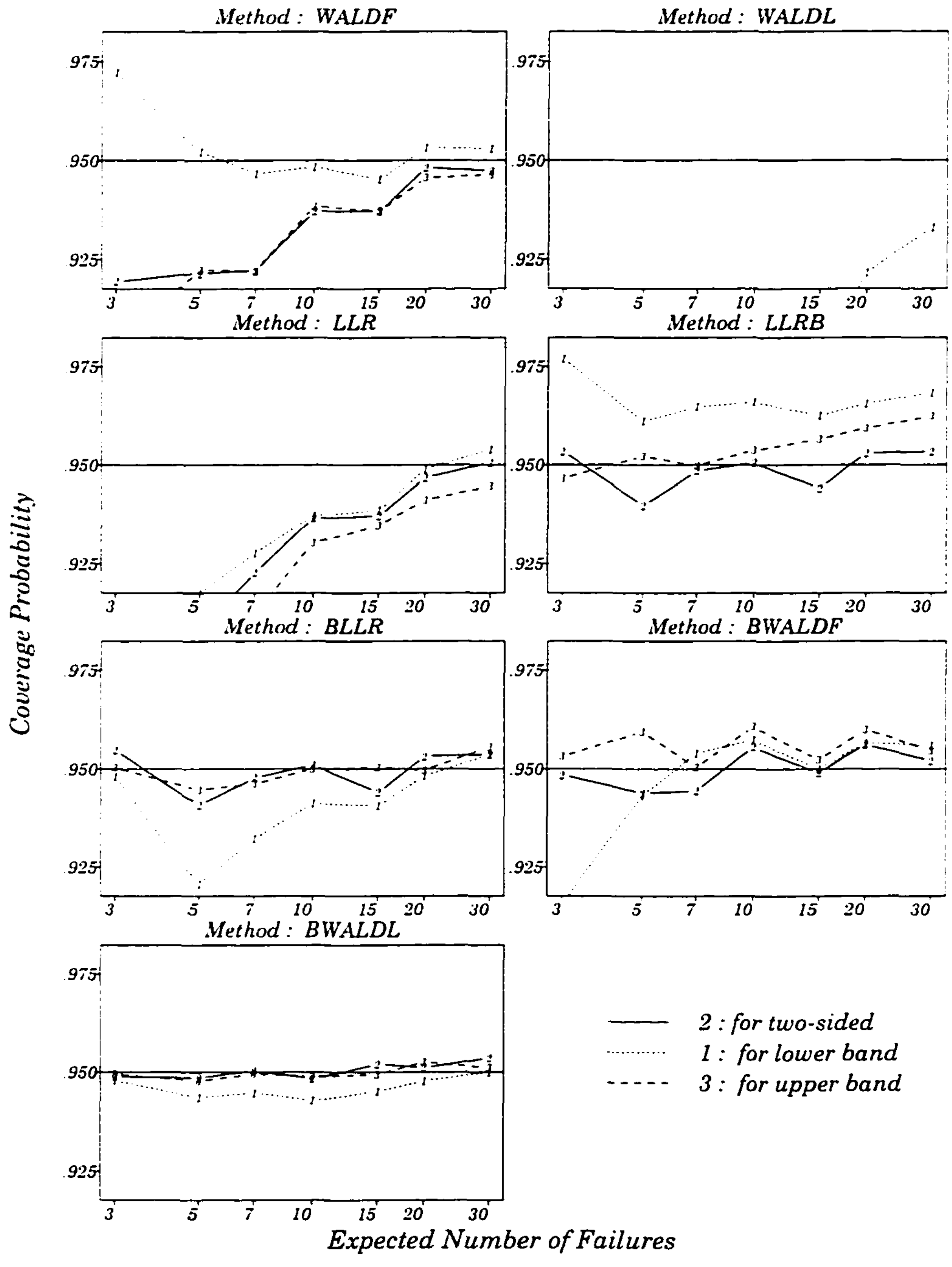



Figure 4.6 Coverage probability plot of methods for constructing approximate 95% confidence regions for two-sided and one-sided simultaneous confidence bands in the cases $p_f = .9$.

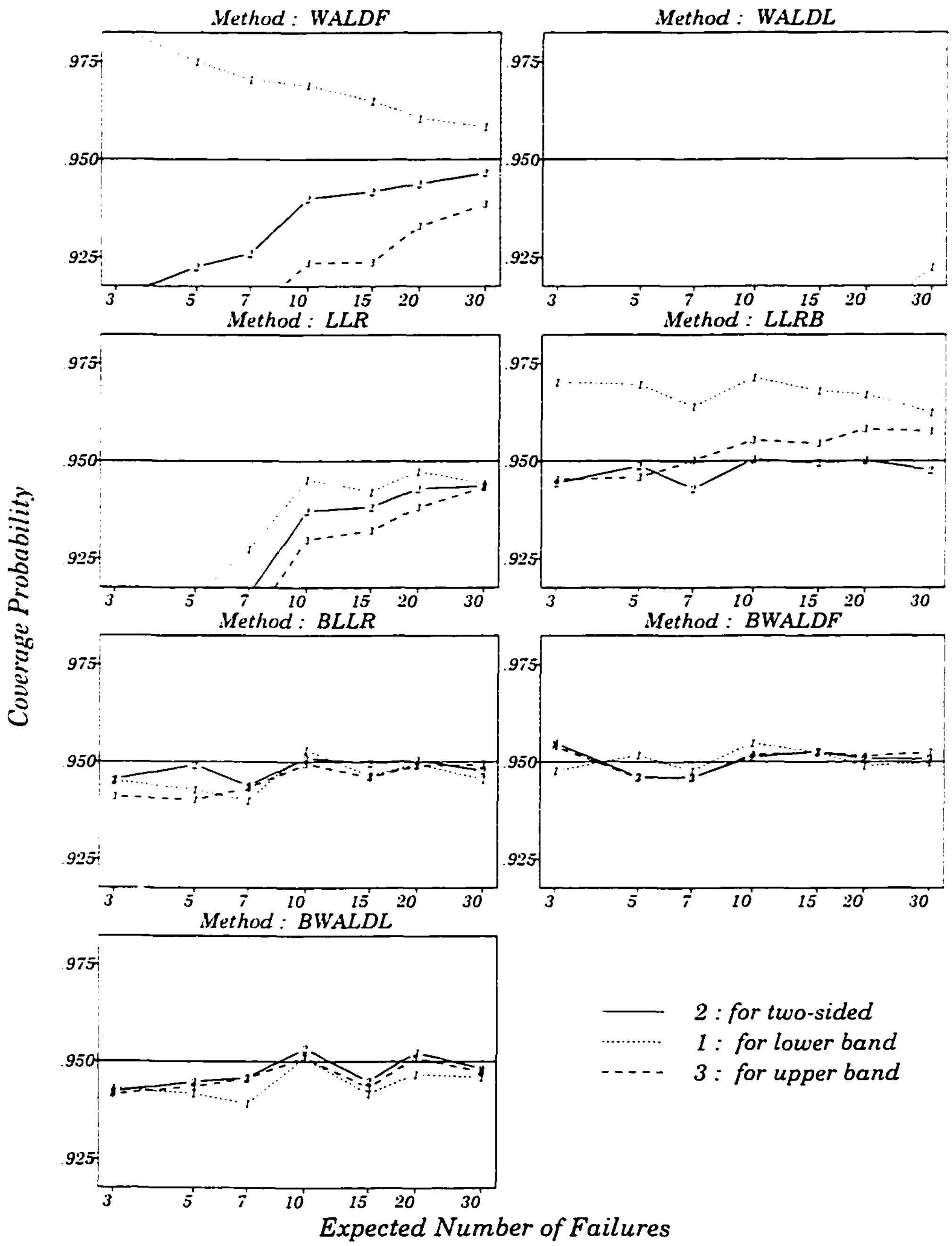



Figure 4.7 Coverage probability plot of the methods for constructing approximate 95% two-sided and one-sided simultaneous confidence bands in the cases $p_f = 1.0$.

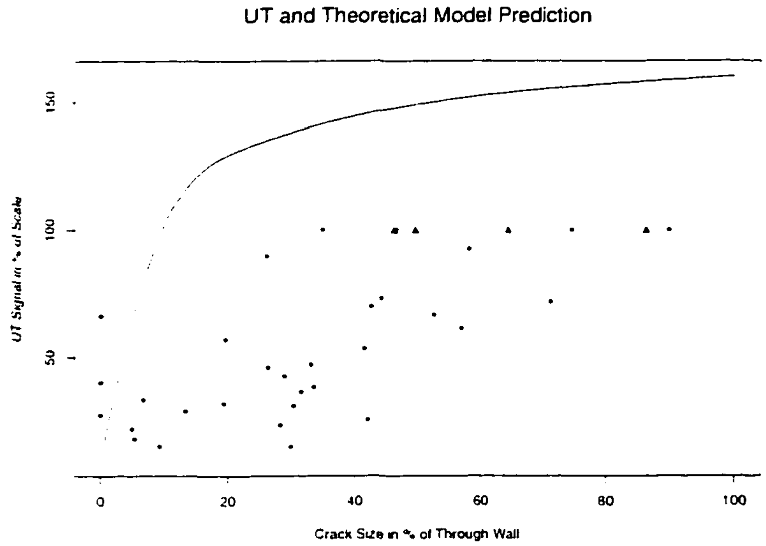



Figure 4.8 The UT signals and the theoretical model predictions. Censored data is represented by triangles.

will saturate when it exceeds a specific bound. The ultrasonic signals were reported in the scaled format as the percentage of a full-scale signal which is determined by calibrating on a given standard. Signals above 100% are right censored. Figure 4.8 shows the observed data and the prediction from a theoretical physical model for ideal flaws (rectangular slots). The signals from the ideal flaws are much stronger than actual cracks. Modeling the deviations (bias and variance) provides a useful model for estimating POD for the inspection method.

Let $\tilde{y}_k$ denote the prediction from the theoretical physical model for ultrasonic NDE signals (UNDE model) for a crack size of $a_k$. We define the generalized deviations (using a Box-Cox transformation) as

$$g(y_k, \tilde{y}_k; \lambda) = \begin{cases} \frac{(y_k)^\lambda - 1}{\lambda} - \frac{(\tilde{y}_k)^\lambda - 1}{\lambda}, & \lambda \neq 0, \\ \log(y_k) - \log(\tilde{y}_k), & \lambda = 0. \end{cases} \tag{4.12}$$

The purpose of using the generalized transformation is to simplify the modeling of variability in the UT signals (specifically to stabilize variance and obtain a simple form for the distribution).

Based on the experiences with large amounts of experimental UT data, Meeker et al. (1995, 1996) observed that a value of $\lambda$ in the neighborhood of 0.3 tends to make the distribution of the deviation close to iid $N(\mu, \sigma^2)$. For the heat exchanger UT data. Sarkar et al. also found that $\lambda = 0.3$ is suitable.

For the heat exchanger UT data, the scaled UT signal amplitude was recored in the form of single right censoring with the fixed right censoring level $t_c$. The generalized deviation results in multiplely right-censored values $x_{ci} = g(t_c, \tilde{y}_i)$, $i = 1, \ldots, n$. We use the method of maximum likelihood to estimate the unknown parameters $\mu$ and $\sigma$. Figure 4.9 shows a normal probability plot and 95% pointwise confidence intervals for the distribution of the generalized deviations. We see that the normal distribution fits the generalized deviation data well.

Let $Y$ be the maximum reading in the gate of an UT A-scan. The threshold $y_{th}$ was chosen to be the 25% of the full-scale signal. There is a detection when $Y > y_{th}$. For this application, the POD is of the primary interest. Under the general model the probability of a detection on any given reading of a crack of size $a$ is

$$\begin{aligned} POD(a) &= \Pr(Y(a) > y_{th}) = 1 - \Pr[g(Y(a), \tilde{y}(a)) \le g(y_{th}, \tilde{y}(a))] \\ &= 1 - \Phi_{\text{nor}}\left[\frac{g(y_{th}, \tilde{y}(a)) - \widehat{\mu}_g}{\widehat{\sigma}_g}\right] \end{aligned} \tag{4.13}$$

where $\Phi_{\text{nor}}$ is the standard normal (Gaussian) cumulative distribution function and $\widehat{\mu}_g$ and $\widehat{\sigma}_g$ are estimates from the generalized deviation data.

Sarkar et al. (1998) provide point-wise confidence intervals for a POD curve by using the delta method and a normal approximation. For actual applications of system reliability, one would be interested in the uncertainty of the estimation of the POD curve for a range of crack sizes. The methods developed in this paper provide the needed simultaneous confidence bands.

The following gives the bootstrap procedure used to find the critical value $\gamma$ for the methods being considered. Let $U(\mu, \sigma)$ be the particular statistic used for finding the confidence region. This statistic could be $Q_F$, $Q_L$ or $W$ that is defined in the Section 4.3.

1. Simulate one sample $x_1^*, \ldots, x_n^*$ from the normal distribution $N(\widehat{\mu}, \widehat{\sigma}^2)$.
2. Let $\delta_i^* = 0$ if $x_i^* \le x_{ci}$ and $\delta_i^* = 1$ if $x_i^* > x_{ci}$. Set $z_i^* = \min\{x_i^*, x_{ci}\}$, $i = 1, \ldots, N$. Calculate $U^*$ from the bootstrap data $(z_i^*, \delta_i^*)$, $i = 1, \ldots, N$.
3. Repeat steps 1 and 2 $B$ times to calculate the MLEs $\widehat{\mu}_j^*$ and $\widehat{\sigma}_j^*$ of $\widehat{\mu}$ and $\widehat{\sigma}$ and $U_j^*$of $U_j$, $j = 1, \ldots, B$. Arrange the $U_j$ in ascending order.
4. Use $U_{B(1-\alpha)}^*$ as the critical value $\gamma$ for the given confidence coefficient $1 - \alpha$ in finding two-sided simultaneous confidence bands.

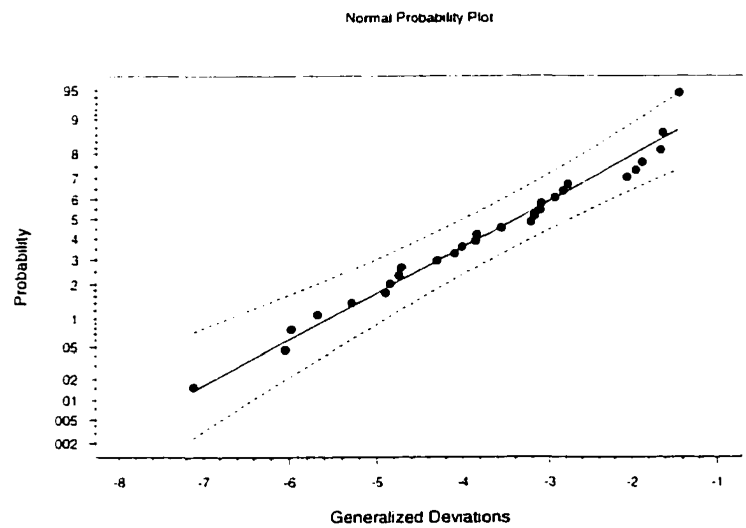



Figure 4.9 Normal probability plot with $\lambda$ =.3 and 95% pointwise confidence interval.

5. Let $k$ be a positive number less than $n$ and let $m = \#\{(\widehat{\mu}_j^*, \widehat{\sigma}_j^*) : j > k, (\widehat{\mu}_j^*, \widehat{\sigma}_j^*) \in S_r'^*\}$ where $S'^*$ is defined in equation 4.11. Find $k$ such that $k + m = B(1 - \alpha)$. Use $U_k^*$ be the critical value $\gamma$ for the given confidence coefficient $1 - \alpha$ in finding one-sided simultaneous confidence bands.

We use $B = 10000$ to calculate the critical value $\gamma$ for the BWALDF and BWALDL methods and to construct the 95% two-sided and 97.5% one-sided simultaneous confidence bands for the POD curve. Figure 4.10 compares the two-sided 95% pointwise confidence intervals using the delta method and the simultaneous confidence bands using the BWALDL method. The important differences are clear. The BWALDL band is wider especially when the crack size is smaller than 20% of referenced size. Figure 4.11 shows that the difference between the BWALDL and the BWALDF methods is not so large. As indicated by the simulation study, because the sample size is 32, the confidence level should be close to the nominal value. Figure 4.12 compares a set of the 97.5% one-sided lower pointwise intervals based on the normal approximation and lower simultaneous confidence bands based on the BWALDF and BWALDL methods. The pointwise intervals tend to lead to narrower region which could be misleading when interest is over a range of crack sizes.

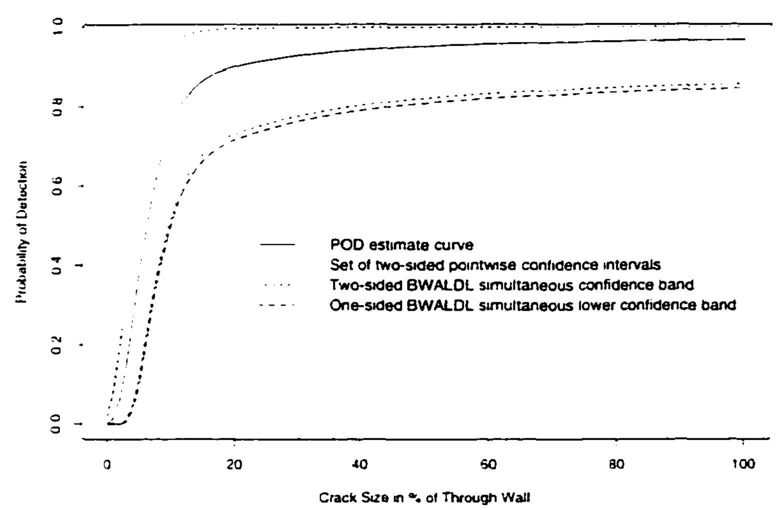



Figure 4.10 The 95% two-sided pointwise confidence intervals using a normal approximation and the 95% two-sided and 97.5% one-sided lower simultaneous confidence bands using the BWALDL method.

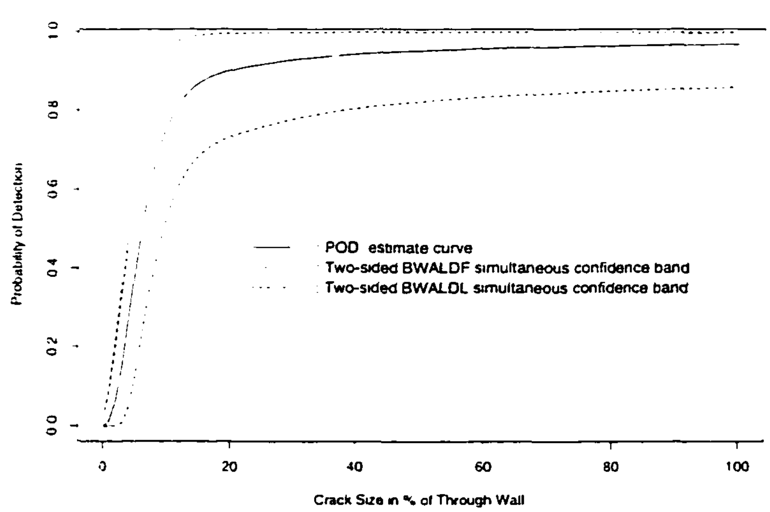



Figure 4.11 The 95% two-sided simultaneous confidence bands calculated by using the BWALDL and BWALDF methods.

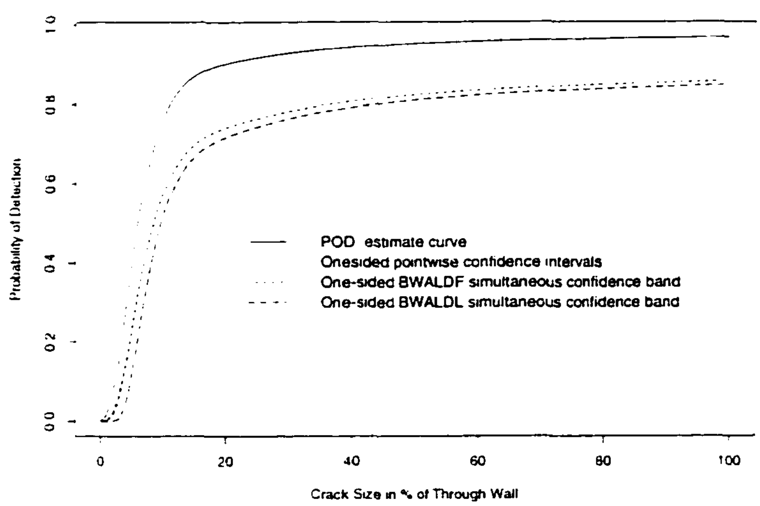



Figure 4.12 The 97.5% one-sided pointwise lower confidence bounds and simultaneous lower confidence bands calculated by using the BWALDL and BWALDF methods.

## 4.6 Discussion and Future Work

Cheng and Iles (1983. 1988) use Wald statistics with (expected) Fisher information and provide different approaches for finding one-sided and two-sided simultaneous confidence bands when there is no censoring in data. We extend their approach by using Wald statistics with local information and likelihood ratio statistics (with or without Bartlett correction) and compare these to corresponding simulation or bootstrap calibrated versions of the same methods when data are complete or censored. The methods presented in this paper can be used to construct two-sided simultaneous confidence bands for general continuous functions. For constructing one-sided simultaneous confidence bands. these methods can only be directly applied to the cdf of location-scale distributions.

We show that for the location-scale model. the accuracy of the procedure for constructing the simultaneous confidence bands is the same as that of the procedure for constructing its corresponding joint confidence region. The BWALDF. BWALDF. and BLLR methods have exact coverage probability when data are complete or Type II censored. When data are Type I censored. only approximate joint confidence regions can be obtained. Our simulation study shows that the BWALDF and BLLR methods provide accurate coverage probabilities when the number of failures reaches 15 for different proportions failing. The BWALDL method produces accurate coverage probabilities when the number of failures reaches 5. The following are some issues for future research:

- In some cases interest centers on inference for a function over some particular range of its arguments. For example, only the lower part of a cdf might be of interest [e.g., $\Phi(x;\theta), x < t$, for some time $t$]. For the cdf of a location-scale distribution, we can construct the corresponding joint confidence regions by following the arguments similar to those in the Sections 4.3.1 and 4.3.2 and then use the resulting regions to construct two-sided and one-sided simultaneous bands, respectively. The shape of the joint confidence region will depend on which part of the function is of interest.

- Both the BWALDF and the BWALDL methods provide accurate joint confidence regions for the unknown parameters in the location-scale model. We use these methods to construct correspondingly accurate simultaneous confidence bands. The open question is how well these two methods perform in other models. In particular, it would be useful to know if they are as good as the BLLR method which generally has second order accuracy in coverage probability for both complete and censored data (Jeng and Meeker 1998). A general method to construct accurate one-sided simultaneous bands for a function $g(\cdot;\theta)$ still needs to be explored. The challenge is to determine an appropriate confidence region that can be used to generate a one-sided simultaneous band.

- The approach used to construct simultaneous confidence bands in this paper can be extended to regression problems. Escobar and Meeker (1998) give a formula to calculate the simultaneous confidence band of a regression curve using the WALDF and WALDL methods for the location-scale model. Their formula can also be used in the BWALDF and BWALDL methods. The simultaneous confidence bands using the LLR, LLRB or BLLR methods also can be obtained numerically.

## References

Beran, R. (1988), "Balanced simultaneous confidence sets," *Journal of the American Statistical Association*, **83**, 679-686.

Cheng, R. C. H. and Iles, T. C. (1983), "Confidence bands for cumulative distribution functions of continuous random variables," *Technometrics*, **25**, 77-86.

Cheng, R. C. H. and Iles, T. C. (1988). "One-sided confidence bands for cumulative distribution functions" *Technometrics*, **30**, 155-159.

Cox and Hinkley (1974). *Theoretical Statistics*. London: Chapman & Hall.

Efron, B., and Tibshirani, R. J. (1993). *An Introduction to the Bootstrap*. New York: Chapman & Hall.

Escobar, L. A. and Meeker, W. Q. (1994). "Fisher information matrix for the extreme value, normal and logistic distributions and censored data". *Applied Statistics*, **43**, 533-540.

Escobar, L. A. and Meeker, W. Q. (1998). "Simultaneous inference with expected and observed information matrix." Preprint, Department of Statistics, Iowa State University.

Hall, P. (1992). *The Bootstrap and Edgeworth Expansion*. New York: Springer-Verlag.

Jeng, S. L. and Meeker, W. Q. (1998). "Bootstrap likelihood ratio statistics". Preprint, Department of Statistics, Iowa State University.

Jensen, J. L. (1993). "A historical sketch and some new results on the improved log likelihood ratio statistic." *Scandinavian Journal of Statistics*, **20**, 1-15.

Kendall, M. and Stuart A. (1991). *The Kendall's Advanced Theory of Statistics*, Fifth Edition, Vol 2. Oxford University Press.

Kanofsky, P. and Srinivasan, R. (1972). "An approach to the construction of parametric confidence bands on cumulative distribution functions." *Biometrika*, **59**, 623-631.

Lawless, J. F. (1982). *Statistical Models and Methods for Lifetime Data*. New York: John Wiley & Sons.

Le Cam, L. (1990). "On the standard asymptotic confidence ellipsoids of Wald." *International Statistical Review*, **58** ,129-152.

Lehmann, E. L. (1986). *Testing Statistical Hypotheses*. New York: John Wiley & Sons.

Meeker, W. Q., Thompson, R. B., Chiou, C-P, Jeng, S-L, and Tucker, W. T. (1996). "Methodology for estimating nondestructive evaluation capability". *Review of Progress in Quantitative NDE* , **15B**, 1983-1991, D.O. Thompson and D.E. Chimenti, Eds., New York: Plenum Press.

Meeker, W. Q., Jeng, S-L, Chiou, C-P. and Thompson, R. B. (1997). "Improved methodology for predicting POD of detecting synthetic hard alpha inclusions in Titanium". *Review of Progress in Quantitative Nondestructive Evaluation*, **17B**, 2021-2028. Thompson, D. O. and Chimenti, D. E., Eds., New York: Plenum Press.

Meeker, W. Q., Jeng, S-L, Chiou, C-P, Thompson, R. B. (1998). "Improved methodology for inspection reliability assessment for detecting synthetic hard alpha inclusions in Titanium". *Review of Progress in Quantitative Nondestructive Evaluation*, Thompson, D. O. and Chimenti, D. E., Eds., **17**, New York: Plenum Press. (in press)

Robinson, J. A. (1983). "Bootstrap confidence intervals in location-scale models with progressive censoring." *Technometrics*, **25**, 179-187.

Rao, C. R. (1973). *Linear Statistical Inference and its Application*, New York: John Wiley & Sons.

Sarkar, P., Meeker, W. Q., Thompson, R. B., Gray, T. A., and Junker, W. (1998). "Probability of detection modeling for ultrasonic testing.", *Review of Progress in Quantitative Nondestructive Evaluation*, Thompson, D. O. and Chimenti, D. E., Eds., **17**, New York: Plenum Press. (in press)

Serfling, R. J. (1980) *Approximation Theorems of Mathematical Statistics*, New York: John Wiley & Sons.

Yeh, A. B. and Singh, K. (1997). "Balanced confidence regions based on Tukey's depth and the bootstrap." *Journal of the Royal Statistical Society, Series B*, **59**, 639-652.

# 5 CONCLUSION

The first paper of our research provides a detailed comparison of methods for constructing approximate confidence intervals. These methods range from the most commonly used large-sample normal-approximation methods to the more modern computationally intensive likelihood and simulation-based methods. Because opposite lower and upper bounds of a two-sided confidence interval tend to have conservative versus anti-conservative coverage probabilities. the effect of averaging often results in reasonably adequate coverage-probability approximations for two-sided confidence intervals in situations with moderately large sample sizes. Our results show. however. that for moderate amounts of censoring and one-sided bounds (most commonly used in practical applications in the physical and engineering sciences as well as other areas of application) the simple normal-approximation (NORM and TNORM) methods provide only crude approximations even when the expected number of failures is as large as 100.

Appropriate computationally-intensive methods provide important improvements. In particular. likelihood-based methods. even when calibrated with the large sample chi-square distribution approximation (e.g.. the LR method). generally provide better results. Calibrating the LR CIs by simulation (see Appendix A) does not address the asymmetry problem and results in inaccurate one-sided bounds. Calibrating the individual tails of a likelihood-based interval with simulation (i.e.. the PBSRLLR method) provides important improvements in coverage probability accuracy. even for small $\mathrm{E}(r)$. for all but one exceptional situation (i.e.. inferences at times near to the censoring time or quantiles near the proportion censoring with $E(r) \leq 10$). The transformed bootstrap-$t$ procedure provides a computationally simpler method. but one needs to be careful in the specification of the transformation to be used.

In the second paper we prove that the distributions of likelihood ratio statistics and their signed square root can be approximated by their bootstrap distribution up to the second order $[O(1/n)]$ when the underling distribution is partly discrete. One application of this result can be used to construct one-sided confidence bounds. two-sided confidence intervals or joint confidence region for complete and censored data.

Some examples like the one-parameter exponential model and logistic regression given by Jensen (1989, 1993) are stated here to illustrate the applications. The two-parameter Weibull distribution model is studied in details when data is Type I censored. Our simulation study compares several commonly suggested methods (Bootstrap-$t$ and $BC_a$) and other higher order accurate methods (modified signed square root likelihood ratio statistics) with likelihood ratio statistics calibrated by bootstrap procedures and provide a clear view of their finite-sample properties. Although the LLRB method for two-sided confidence intervals is the most accurate one in coverage probability, the resulting two-sided confidence interval is not symmetric in the sense that the confidence level of one side of the interval is larger than the nominal confidence level and the confidence level of the other side of the interval is smaller than the nominal one. If one-sided confidence bounds are of interest, the PBSRLLR and PBMSRLLR methods provide better coverage probability when the expected number of failures exceeds 10.

In the third paper we focus on the problem of computing simultaneous confidence bands, we provide methods by using the Wald statistics with local information and the likelihood ratio statistics (with or without a Bartlett correction) and compare these to corresponding simulation or bootstrap calibrated versions of the same methods when data is complete or censored. The methods presented can be used to construct two-sided simultaneous confidence bands for general continuous functions. For constructing one-sided simultaneous confidence bands, these methods can only be applied directly to cdf of location-scale distributions.

The accuracy of the resulting simultaneous confidence bands depend on the coverage probability of its corresponding joint confidence region. We show that in location-scale model, the BWALDF, BWALDF, and BLLR methods can be used to construct joint confidence regions with exact coverage probability when data is complete or Type II censored. When data is Type I censored, only approximate joint confidence regions can be obtained. Our simulation study shows that the BWALDF, BWALDF, and BLLR methods provide accurate coverage probability when number of failure reaches 15 for different proportion failings. The BWALDL method produces accurate coverage probability when number of failure reaches 5.

In addition to providing guidance for practical applications, our results suggest the following avenues for further research.

1. Our study leaves unanswered the question of what one should do when making inferences in the exceptional case when the failure number is down to 10. We see no easy solution to this problem. Some possibilities include

- Find a smoothed bootstrap distribution of MLE when sample size is small (less than 10). Another alternative suggested by some limited simulation results is to use a double bootstrap calibration.
- Extending the censoring time of the life test to be safely and sufficiently beyond the time point (or proportion failing) of interest. This requires prior knowledge of the failure-time distribution which is not generally available.
- Design life test experiments to result in Type II censored data. In this case. exact confidence interval procedures are available. but experimenters generally have to deal with time constraints in life testing and thus there may be resistance to such life test plans. On the other hand. Type II censoring provides important control over the amount of information that a life-test experiment will provide.
- Design life test experiments to result in multiple time-censoring (where the results of Robinson (1983) suggest that excellent large sample approximations are available from computationally intensive methods). In this case. constraints on time or number of units available for testing may also lead to resistance to such life test plans.
- If none of the above is possible (e.g.. for reasons given above or because the experiment has already been completed) it might be possible to make use of nonparametric methods (where conservative confidence intervals or bounds may be available if there is a sufficient amount of data).

2. Our study has focused on the Weibull distribution. It would be of interest to replicate the study for other distributions. We would expect very similar results for other log-locations-scale distributions such as the lognormal and the loglogistic distributions.

3. It would be of interest to extend this study to other censored-data situations that arise in applications. including regression analysis and the analysis of accelerated life test data. more complicated censoring schemes like interval censoring and random censoring. simultaneous confidence interval and bounds. intervals to compare two different grouped. and so on.

4. Our examples show that the theorems we have developed can be applied to logistic regression and location-scale model with Type I censoring data. For other kinds of

censoring and distribution. Conditions (A) and (B) in Section 3.2 can be expected to hold when the model distribution is smooth and without overly heavy tails.

5. The modified likelihood ratio statistic [presented by Barndorff-Nelson (1986, 1991)] seems to provide better coverage probabilities when they are calibrated with a bootstrap procedure. The order of accuracy could be further explored.

6. Both BWALDF and BWALDL methods provide accurate joint confidence regions for the unknown parameters. We use these to provide correspondingly accurate simultaneous confidence bands. The open question is how well these two methods perform in other models. Will they still be as good as the BLLR method which generally has second order accuracy in coverage probability for both complete and censored data?

7. A general method to construct one-sided simultaneous bands for a function of parameters still need to be explored. The challenge is to determine an appropriate confidence region corresponding to the desired one-sided simultaneous band.

8. In some cases interest centers on inference for a function over some particular range of its arguments. For example, when the lower half of cdf is of interest [i.e., $\Phi(x;\theta), x < t$, for some time $t$], for the cdf of a location-scale distribution, we can construct the corresponding joint confidence regions by following arguments similar to those in Sections 4.3.1 and 4.3.2 for two-sided and one-sided simultaneous bands. The shape of the joint confidence region depends on which part of the function is of interest. For general functions, more research is needed.

# APPENDIX A CALCULATION DETAILS FOR FINDING MLE AND CONFIDENCE INTERVALS

## Calculation of MLE and CIs

### Calculation of ML Estimator

Consider a sample of $n$ observations with $t_1, \ldots, t_r$ reported as exact failure times (suppose that $2 \leq r \leq n$) and $n - r$ observations censored at a common time $t_c$. A simple expression can be derived for the Weibull log likelihood in the $\eta = \exp(\mu)$ and $\beta = 1/\sigma$ parameterization. The Weibull ML estimates can be obtained by solving the following two equations:

$$\left[\frac{\sum_{i=1}^{r} t_i^{\beta}\log(t_i) + (n-r)t_c^{\beta}\log(t_c)}{\sum_{i=1}^{r} t_i^{\beta} + (n-r)t_c^{\beta}} - \frac{1}{\beta}\right] - \frac{1}{r}\sum_{i=1}^{r}\log(t_i) = 0$$

$$\eta^{\beta} = \frac{1}{r}\left[\sum_{i=1}^{r} t_i^{\beta} + (n-r)t_c^{\beta}\right].$$

Note that the first equation does not contain $\eta$ and is thus easy to solve numerically for $\beta$. For more detail see Lawless (1982, page 170).

### Calculation of CI from log likelihood ratio (LLR) Method

Suppose $C = \{\tilde{\theta}_1 : \text{where } (\tilde{\theta}_1, \tilde{\theta}_2) \text{ are the solutions of the following two equations.}\}$

$$\frac{\partial}{\partial\theta_2}\left[\log\left(\frac{L(\theta_1, \theta_2)}{L(\widehat{\boldsymbol{\theta}})}\right)\right] = 0 \tag{A.2.1}$$

$$-2\log\left(\frac{L(\theta_1, \theta_2)}{L(\widehat{\boldsymbol{\theta}})}\right) = \chi^2_{(1-\alpha, 1)} \tag{A.2.2}$$

Then the $100(1-\alpha)\%$ CI of $\theta_1$ is $(\min_{\tilde{\theta}_1 \in C}\tilde{\theta}_1, \max_{\tilde{\theta}_1 \in C}\tilde{\theta}_1)$

**Calculation of CI from log likelihood ratio Bartlett corrected (LLRBART) Method**

Suppose $C' = \{\check{\theta}_1$ : where $(\check{\theta}_1, \check{\theta}_2)$ are the solutions of the equation (A.2.1) and the following equation.$\}$

$$-2\log\left(\frac{L(\theta_1, \theta_2)}{L(\widehat{\boldsymbol{\theta}})}\right)/\mathrm{E}(v)^* = \chi^2_{(1-\alpha,1)},$$

where $E(v)^* = \sum -2\log\max_{\theta_2}[L_i^*(\widehat{\theta}_1, \theta_2)/L_i^*(\widehat{\boldsymbol{\theta}}_i^*)]/2000$, which is the bootstrap estimate of expected value of $v$. The $100(1-\alpha)\%$ CI of $\theta_1$ is $(\min_{\check{\theta}_1\in C'}\check{\theta}_1, \max_{\check{\theta}_1\in C'}\check{\theta}_1)$.

**Calculation of CI from parametric bootstrap signed-root log-likelihood ratio (PBSRLLR) Method**

The lower $100(1-\alpha/2)\%$ confidence limit of $\theta_1$ is the solution of the equation (A.2.1) and the following equation.

$$\mathrm{sign}(\widehat{\theta}_1 - \theta_1)\{-2\log[L(\theta_1, \theta_2)/L(\widehat{\boldsymbol{\theta}})]\}^{1/2} = z_{\widehat{\theta}^*_{1(\alpha/2)}},$$

where $z_{\widehat{\theta}^*_{1(\alpha/2)}}$ is the $\alpha/2$ quantile of the distribution of $\mathrm{sign}(\widehat{\theta}_1^* - \widehat{\theta}_1)\{-2\log\max_{\theta_2}[L^*(\widehat{\theta}_1, \theta_2)/L^*(\widehat{\boldsymbol{\theta}}^*)]\}^{1/2}$.

## Other Methods Considered

The graphs used to present the results of our simulation contained comparisons among only a subset of the confidence interval methods that we compared. Dropping these from our detailed comparisons did not affect our primary conclusions. We mention the other methods for completeness.

**Parametric bootstrap of LR method (PBLLR)** Let $R^*$ be the bootstrap version of the profile likelihood statistic R (Equation 2.1 in Section 2.3.2). Suppose $z_{\widehat{\theta}^*_{1(\alpha)}}$ is the $\alpha$ quantile of the distribution of $R^*$. The two-sided $100(1-\alpha)\%$ confidence interval is obtained from $\min\{R^{-1}(z_{\widehat{\theta}^*_{1(\alpha)}})\}$ and $\max\{R^{-1}(z_{\widehat{\theta}^*_{1(\alpha)}})\}$.

**Robinson's Method** Robinson (1983) provide a parametric bootstrap method to any model that is transformable to location-scale form and has invariant estimators. Let $(\widehat{\mu}, \widehat{\sigma})$ be the ML estimator of $(\mu, \sigma)$ and $\widehat{t}_p$ is defined as in Section 2.1. Let $z_{\widehat{\sigma}^*_{(\alpha)}}$ and $z_{(\widehat{\mu}^*, \widehat{\sigma}^*)_{(\alpha)}}$ be the $\alpha$ quantile of the distribution of $\widehat{\sigma}/\widehat{\sigma}^*$ and $[\log(\widehat{t}_p^*) - \log(\widehat{t}_p)]/\widehat{\sigma}$.

Then confidence intervals for $\sigma$ and $t_p$ can be calculated form $(\widehat{\sigma}^* z_{\widehat{\sigma}_{(\alpha/2)}}, \widehat{\sigma} z_{\widehat{\sigma}^*_{(1-\alpha/2)}})$ and $(\exp(\log(\widehat{t_p}) - \widehat{\sigma} z_{(\widehat{\mu}^*, \widehat{\sigma}^*)_{(1-\alpha/2)}}), \exp(\log(\widehat{t_p}) - \widehat{\sigma} z_{(\widehat{\mu}^*, \widehat{\sigma}^*)_{(\alpha/2)}}))$.

# APPENDIX B THEOREMS AND METHODS USED

## Theorems Used in the Paper

We want to obtain an expansion for the distribution of the statistic $\sqrt{n}g(S_n/n)$ where $S_n = (X_n, Y_n)$. $X_n$ is a continuous variable with means zero in $\mathbb{R}^{q_1}$. $Y_n$ is a lattice variable mean $\mu_n$ in $\mathbb{R}^{q_2}$ having minimal lattice $\mathbb{Z}^{q_2}$ and. and $g$ is a smooth function. Let $q = q_1 + q_2$. We consider only the case $q_1 > q_2$. For a function of $t \in \mathbb{N}^k$ we denote by $\partial^\nu$ the partial derivative $\partial^{|\nu|}/(\partial t_1^{\nu_1} \ldots \partial t_k^{\nu_k})$ where $\nu \in \mathbb{R}^k$. $|\nu| = \sum \nu_i$ and $\nu! = \nu_1! \ldots \nu_k!$. To formulate Lemma 1 we let $\Sigma$ denote a $k \times k$ positive definitive matrix. The multivariate normal cdf with mean vector 0 and covariance $\Sigma$ is denoted by $\Phi_\Sigma$ and the corresponding density by $\phi_\Sigma$.

The expansion for $\sqrt{n}g(S_n/n)$ is formulated in terms of Borel sets satisfying certain conditions on their boundary. The $\delta$-boundary of a Borel set $A$ is defined as

$$(\partial A)^\delta = \{B(x,\delta) | x \in A, B(x,\delta) \not\subseteq A\}.$$

where $B(x,\delta)$ denotes a sphere centered at $x$ and with radius $\delta$.

The following lemma is the same as Theorem 1 of Jensen (1989). It establishes the Edgeworth expansion for the statistic $\sqrt{n}g(S_n/n)$ under fairly general conditions.

**Lemma 5 (Jensen, 1989)** *Suppose that the following assumptions are satisfied.*

*(i) The r-th moment of $S_n$ is finite, i.e. $\mathrm{E}\| S_n \|^r < \infty$, where $r = \max\{2s-1, q_1+1\}$ with $s \geq 3$.*

*(ii) There exist $\varepsilon > 0$ and $n_0 \in \mathbb{N}$ such that for $n \geq n_0$ and $\| (t,v) \| \leq \varepsilon$*

$$f_n(t,v) = \mathrm{E}(e^{itX_n + iv(Y_n - \mu_n)}) = e^{nH_n(t,v)}$$

*for some function $H_n(t,v)$.*

*(iii)* $\partial^\nu H_n(t,v)|_0 \to \kappa_\nu$ *as* $n \to \infty$ *and* $\nu \in \mathbb{N}^{q_1+q_2}$, $|\nu| \le r$, *where* $\kappa_\nu$ *is finite and the matrix of second partial derivatives* $\{\kappa_\nu\}_{|\nu|=2}$ *is negative definite. Furthermore, for* $n \ge n_0$ *and* $\| (t,v) \| \le \varepsilon$ *the r-th order partial derivative* $\partial^\nu H_n(t,v), |\nu| = r$, *is bounded.*

*(iv) For any* $\varepsilon > 0$ *there exist* $c > 0$ *and* $\rho < 1$ *such that* $|\partial^\nu f_n(t,v)| < c\rho^n$ *for* $|\nu| \le r, \| t \| > \varepsilon$ *for all* $v$.

*(v) For any* $\varepsilon_1 > 0$ *there exist* $\varepsilon, c > 0$, *and* $\rho < 1$ *such that* $|\partial^\nu f_n(t,v)| < c\rho^n$ *for* $|\nu| \le r, \| v \| > \varepsilon_1, |v_i| \le \pi$ *and* $\| t \| < \varepsilon$.

*(vi)* $g : \mathbb{R}^{q_1+q_2} \to \mathbb{R}^k$ *with* $k \le q_1$ *has derivatives of order* $s-1$ *which are continuous in a neighborhood of* $(0, \mu_n/n)$. *Also* $g(0, \mu_n/n) = 0$.

*(vii) The* $q_1 \times k$ *matrix*

$$D_0 = \left.\frac{\partial g(x,y)}{\partial x'}\right|_{(0,\mu_n/n)}$$

*has full rank.*

*Then we have that for any* $d > 0$ *there exists a constant* $c$ *such that*

$$\left| \Pr(\sqrt{n} g(X_n/n, Y_n/n) \in A) - \int_A \xi_{s,n}(z)dz \right| \le cn^{-(s-2)/2}$$

*for all Borel sets* $A$ *with* $\Phi_{\Sigma_0}[(\partial A)^\delta] \le d\delta$ *for every* $\delta > 0$. *As defined in Bhattacharya and Ghosh (1978),* $\int_A \xi_{s,n}(z)dz$ *is the formal Edgeworth expansion of the distribution function of* $\sqrt{n} g(X_n/n, Y_n/n)$, *where*

$$\xi_{s,n}(z) = \phi_{\Sigma_0}(z) \left[ 1 + \sum_{j=1}^{s-2} n^{-j/2} Q_j(z) \right] \tag{B.1}$$

*and* $Q_j$ *is a polynomial in* $z$. *Here* $\Sigma_0$ *is positive definite and the coefficients of* $Q_j$ *are of order one* $[O(1)]$.

**Proof.** See Jensen (1989) Section 2. ■

**Remark 2** *From Remark 1.4 of Bhattacharya and Ghosh (1978),* $Q_j$ *depends only on the moments of* $S_n$. *When the distribution function of* $S_n$ *is continuous in the unknown parameter* $\theta$, $Q_j$ *will be a continuous function of* $\theta$.

A case of much practical importance is when $g$ is one dimensional. Let the cumulants of order $m$ of $S_n/n = (X_n/n, Y_n/n)$ be denoted by $\lambda_{j_1\cdots j_m}$. That is.

$$\lambda_{j_1\cdots j_m} = \frac{1}{i^m}\frac{\partial^m \log[\mathrm{E}(e^{itS_n/n})]}{\partial t_{j_1}\ldots\partial t_{j_m}}\bigg|_0$$

with $j_1,\ldots,j_m \in \{1,\ldots,q\}$. Define

$$g_{j_1\cdots j_m} = \frac{\partial^r g[\mu + \mathrm{E}(S_n)/n]}{\partial\mu_{j_1}\ldots\partial\mu_{j_r}}\bigg|_0.$$

Sargon (1976) gives the Edgeworth expansion of the statistics $\sqrt{n}g(S_n/n)$ when the underlying distribution of $S_n$ is continuous and $g$ is a real value function. The following Corollary gives a similar result that the underlying distribution of $S_n$ could be partly discrete.

**Corollary 1** *Let $g$ be a one dimensional function and let $s = 5$. Under the assumption of Lemma 5, the expansion to approximate the probability* $\Pr[\sqrt{n}g(S_n/n) \le w]$ *becomes*

$$\Phi_{\sigma^2}(w) + \phi_{\sigma^2}(w)\Bigg\{\frac{-1}{\sqrt{n}}\left[\frac{1}{2}\alpha_4 + \frac{\alpha_1+3\alpha_3}{6\sigma^2}\left(\left(\frac{w}{\sigma}\right)^2 - 1\right)\right]$$
$$+\frac{1}{n}\Bigg[-\frac{\alpha_5+\alpha_7+\alpha_9/2+\alpha_4^2/4}{2\sigma}\frac{w}{\sigma} + \frac{\alpha_2+12\alpha_{10}+4\alpha_6+2\alpha_1\alpha_4+6\alpha_3\alpha_4+12\alpha_8}{24\sigma^3}$$
$$\left(3\frac{w}{\sigma} - \left(\frac{w}{\sigma}\right)^3\right) + \frac{\alpha_1^2+6\alpha_1\alpha_3+9\alpha_3^2}{72\sigma^5}\left(-\left(\frac{w}{\sigma}\right)^5 + 10\left(\frac{w}{\sigma}\right)^3 - 15\frac{w}{\sigma}\right)\Bigg]\Bigg\} + o\left(\frac{1}{n}\right). \tag{B.2}$$

*Here the coefficients are given by*

$$\sigma^2 = \lambda_{jk}g_jg_k,\quad \gamma_a = \lambda_{aj}g_j,\quad \beta_a = \lambda_{ajk}g_jg_k,\quad \gamma_{ab} = \lambda_{abj}g_j,$$
$$\alpha_1 = \lambda_{jkm}g_jg_kg_m,\quad \alpha_2 = \lambda_{jkmq}g_jg_kg_mg_q,\quad \alpha_3 = \gamma_ag_{ab}\gamma_b,$$
$$\alpha_4 = g_{ab}\lambda_{ab},\quad \alpha_5 = g_{ab}\gamma_{ab},\quad \alpha_6 = g_{abc}\gamma_a\gamma_b\gamma_c,$$
$$\alpha_7 = g_{abc}\lambda_{ab}\gamma_c,\quad \alpha_8 = \gamma_ag_{ab}\lambda_{bc}g_{cd}\gamma_d,$$
$$\alpha_9 = \lambda_{ad}g_{dc}\lambda_{cb}g_{ba},\quad \alpha_{10} = \gamma_ag_{ab}\beta_b.$$

*The sum over the subscript is omitted for abbreviation.*

**Proof.** The result follows from Lemma 5 and the result in Sargon (1976). ■

**Remark 3** *(B.2) can be represented as*

$$\Phi_{\sigma^2}(w) - \frac{1}{\sqrt{n}}a_1(w)\phi_{\sigma^2}(w) + \frac{1}{n}a_1(w)\phi_{\sigma^2}(w) + o\left(\frac{1}{n}\right). \tag{B.3}$$

*where $a_1(w)$ and $a_2(w)$ are polynomials with degree 2 and 5, respectively. Note that $a_1(w)\phi_{\sigma^2}(w)$ and $a_2(w)\phi_{\sigma^2}(w)$ are bounded over $w$. If $s = 4$ in Corollary B, the expansion to approximate the probability* $\Pr[\sqrt{n}g(S_n/n) \leq w]$ *becomes*

$$\Phi_{\sigma^2}(w) - \frac{1}{\sqrt{n}}a_1(w)\phi_{\sigma^2}(w) + O\left(\frac{1}{n}\right). \tag{B.4}$$

## Checking Regularity Conditions

We would like to check that the (A) Conditions hold for Equation (3.32). It is clear that if $\Phi$ has a $\nu$-th derivative w.r.t. $\theta$ on $\mathfrak{X} \times \mathfrak{A}$, then Condition (A1) holds. This is true for the SEV, normal, and logistic distributions.

For condition A2, we present the general formulation for location-scale distributions and then discuss the details for the SEV, normal, and logistic distributions.

Let $\xi_i = (x_i - x_p)/\sigma + u_p$, then

$$\begin{aligned}
&\frac{\partial \xi_i}{\partial x_p} = -\frac{1}{\sigma}, \quad \frac{\partial^j \xi_i}{\partial x_p^j} = 0, \quad j \geq 2, \\
&\frac{\partial \xi_i}{\partial \sigma} = -\frac{x_i - x_p}{\sigma^2}, \quad \frac{\partial^j \xi_i}{\partial \sigma^j} = (-1)^j \frac{x_i - x_p}{\sigma^{j+1}}, \quad j \geq 2, \\
&\frac{\partial^{j+k} \xi_i}{\partial x_p^j \partial \sigma^k} = 0, \quad j \geq 2, k \geq 1, \quad \frac{\partial^{j+k} \xi_i}{\partial x_p^j \partial \sigma^k} = \left(-\frac{1}{\sigma}\right)^{k+1}, \quad j = 1, k \geq 1.
\end{aligned} \tag{B.5}$$

Let $\xi_c = (x_c - x_p)/\sigma + u_p$ denote the standardized censoring time, then

$$\begin{aligned}
&\frac{\partial \xi_c}{\partial x_p} = -\frac{1}{\sigma}, \quad \frac{\partial^j \xi_c}{\partial x_p^j} = 0, \quad j \geq 2, \\
&\frac{\partial \xi_c}{\partial \sigma} = -\frac{x_c - x_p}{\sigma^2}, \quad \frac{\partial^j \xi_c}{\partial \sigma^j} = (-1)^j \frac{x_c - x_p}{\sigma^{j+1}}, \quad j \geq 2, \\
&\frac{\partial^{j+k} \xi_c}{\partial x_p^j \partial \sigma^k} = 0, \quad j \geq 2, k \geq 1, \quad \frac{\partial^{j+k} \xi_c}{\partial x_p^j \partial \sigma^k} = \left(-\frac{1}{\sigma}\right)^{k+1}, \quad j = 1, k \geq 1.
\end{aligned} \tag{B.6}$$

The partial derivatives of log likelihood function are

$$\partial^{(0,1)}l(x_i;(\sigma,x_p)) = \delta_i\left[\frac{\phi'(\xi_i)\frac{\partial\xi_i}{\partial x_p}}{\phi(\xi_i)}\right] + (1-\delta_i)\left[-\frac{\phi(\xi_c)\frac{\partial\xi_c}{\partial x_p}}{1-\Phi(\xi_c)}\right],$$

$$\partial^{(1,0)}l(x_i;(\sigma,x_p)) = \delta_i\left[-\frac{1}{\sigma}+\frac{\phi'(\xi_i)\frac{\partial\xi_i}{\partial\sigma}}{\phi(\xi_i)}\right] + (1-\delta_i)\left[-\frac{\phi(\xi_c)\frac{\partial\xi_c}{\partial\sigma}}{1-\Phi(\xi_c)}\right],$$

$$\partial^{(1,1)}l(x_i;(\sigma,x_p)) = \delta_i\left\{\frac{\phi(\xi_i)[\phi''(\xi_i)\frac{\partial\xi_i}{\partial x_p}\frac{\partial\xi_i}{\partial\sigma}+\phi'(\xi_i)\frac{\partial^2\xi_i}{\partial x_p\partial\sigma}]-[\phi'(\xi_i)]^2\frac{\partial\xi_i}{\partial x_p}\frac{\partial\xi_i}{\partial\sigma}}{\phi(\xi_i)^2}\right\}$$
$$+(1-\delta_i)\left\{\frac{[1-\Phi(\xi_c)][-\phi'(\xi_c)\frac{\partial\xi_c}{\partial x_p}\frac{\partial\xi_c}{\partial\sigma}+\phi(\xi_c)\frac{\partial^2\xi_c}{\partial x_p\partial\sigma}]-[\phi(\xi_c)]^2\frac{\partial\xi_c}{\partial x_p}\frac{\partial\xi_c}{\partial\sigma}}{[1-\Phi(\xi_c)]^2}\right\},$$

$$\partial^{(0,2)}l(x_i;(\sigma,x_p)) = \delta_i\left\{\frac{\phi(\xi_i)[\phi''(\xi_i)(\frac{\partial\xi_i}{\partial x_p})^2]-[\phi'(\xi_i)]^2(\frac{\partial\xi_i}{\partial x_p})^2}{\phi(\xi_i)^2}\right\} \tag{B.7}$$
$$+(1-\delta_i)\left\{\frac{[1-\Phi(\xi_c)][\phi'(\xi_c)(\frac{\partial\xi_c}{\partial x_p})^2]-[\phi(\xi_c)]^2(\frac{\partial\xi_c}{\partial x_p})^2}{[1-\Phi(\xi_c)]^2}\right\},$$

$$\partial^{(2,0)}l(x_i;(\sigma,x_p)) = \delta_i\left\{\frac{\phi(\xi_i)[\phi''(\xi_i)(\frac{\partial\xi_i}{\partial\sigma})^2]-[\phi'(\xi_i)]^2(\frac{\partial\xi_i}{\partial\sigma})^2}{\phi(\xi_i)^2}\right\}$$
$$+(1-\delta_i)\left\{\frac{[1-\Phi(\xi_c)][\phi'(\xi_c)(\frac{\partial\xi_c}{\partial\sigma})^2]-[\phi(\xi_c)]^2(\frac{\partial\xi_c}{\partial\sigma})^2}{[1-\Phi(\xi_c)]^2}\right\}.$$

Similarly, for $3 \le |\nu| \ge 4$, $\partial^\nu l(x_i;(\sigma,x_p))$ is a function of $\Phi,\phi,\phi',\phi'',\phi''',\phi''''$, and terms in the Equation (B.5), (B.6).

**SEV Distribution.** For the smallest extreme value distribution,

$$\Phi(\xi) = 1-\exp([-\exp(\xi)]), \quad \phi(\xi) = \exp\{[\xi-\exp(\xi)]\}, \quad \phi'(\xi) = [1-\exp(\xi)]\phi(\xi).$$

Then Equations (B.7):

$$\partial^{(0,1)}l_i = \delta_i\left[(1-\exp(\xi_i))\frac{\partial\xi_i}{\partial x_p}\right] + (1-\delta_i)\left[-\exp(\xi_c)\frac{\partial\xi_c}{\partial x_p}\right],$$
$$\partial^{(1,0)}l_i = \delta_i\left[-\frac{1}{\sigma}+(1-\exp(\xi_i))\frac{\partial\xi_i}{\partial\sigma}\right] + (1-\delta_i)\left[-\exp(\xi_c)\frac{\partial\xi_c}{\partial\sigma}\right]. \tag{B.8}$$

From Equations (B.5), (B.6), and (B.8), $\partial^{(0,1)}l$ and $\partial^{(1,0)}l$ are linearly independent. We can see that all other partial derivatives $\partial^\nu l$, $1\le|\nu|$, are functions of $\exp(\xi_i)$ and the terms in Equations (B.5), (B.6). Then $\tilde{Z}_i$ can be written as

$$(\partial^{(0,1)}l_i,\ \partial^{(1,0)}l_i,\ \delta_i\xi_i\exp(\xi_i),\ \delta_i\xi_i^2\exp(\xi_i),\ \delta_i\xi_i^3\exp(\xi_i),\ \delta_i\xi_i^4\exp(\xi_i),\ \delta_i).$$

Because the expectations of $\xi_i^j \exp(\xi_i)^k$, $0 < j + k$, $0 \leq k \leq 1$, are finite over an open set containing the true parameters, the expectations of $\partial^\nu l$, $1 \leq |\nu|$, are finite over the same open set. This establishes Condition (A2).

**Normal Distribution**. For the normal distribution

$$\phi(\xi) = \frac{1}{\sqrt{\pi}} e^{-\xi^2/2}, \quad \Phi(\xi) = \int_{-\infty}^{\xi} \phi(x)dx, \quad \phi'(\xi) = -\xi\phi(\xi).$$

Then Equations (B.7) becomes:

$$\begin{aligned} \partial^{(0,1)} l_i &= \delta_i \left[\xi_i \frac{\partial \xi_i}{\partial x_p}\right] + (1 - \delta_i) \left[-\frac{\phi(\xi_c)\frac{\partial \xi_c}{\partial x_p}}{1 - \Phi(\xi_c)}\right] \\ \partial^{(1,0)} l_i &= \delta_i \left[-\frac{1}{\sigma} + \xi_i \frac{\partial \xi_i}{\partial \sigma}\right] + (1 - \delta_i) \left[-\frac{\phi(\xi_c)\frac{\partial \xi_c}{\partial \sigma}}{1 - \Phi(\xi_c)}\right]. \end{aligned} \qquad (B.9)$$

Equations (B.5), (B.6) and (B.9), $\partial^{(0,1)} l$, and $\partial^{(1,0)} l$ are linearly independent. We can see that all of the other partial derivatives $\partial^\nu l$, $1 \leq |\nu|$, are functions of $\xi_i$, and terms in (B.5) and (B.6). Then $\tilde{Z}_i$ can be written as

$$\left(\partial^{(0,1)} l_i, \partial^{(1,0)} l_i, \delta_i \xi_i^3, \delta_i \xi_i^4, \delta_i\right).$$

Because the expectations of $\xi_i^j$, $0 < j$, are finite over an open set containing the true parameters, the expectations of $\partial^\nu l$, $1 \leq |\nu|$, are finite over the same open set. This establishes Condition (A2).

**Logistic Distribution**. For the logistic distribution

$$\Phi(\xi) = \frac{1}{1 + e^{-\xi}}, \quad \phi(\xi) = \frac{e^{-\xi}}{(1 + e^{-\xi})^2}, \quad \phi'(\xi) = -\Phi(\xi)\phi(\xi)$$

Then from Equation (B.7):

$$\begin{aligned} \partial^{(0,1)} l_i &= \delta_i \left[-\Phi(\xi_i) \frac{\partial \xi_i}{\partial x_p}\right] + (1 - \delta_i) \left[-\Phi(\xi_c) \frac{\partial \xi_c}{\partial x_p}\right], \\ \partial^{(1,0)} l_i &= \delta_i \left[-\frac{1}{\sigma} - \Phi(\xi_i) \frac{\partial \xi_i}{\partial \sigma}\right] + (1 - \delta_i) \left[-\Phi(\xi_c) \frac{\partial \xi_c}{\partial \sigma}\right]. \end{aligned} \qquad (B.10)$$

From Equation (B.5), (B.6) and (B.10), $\partial^{(0,1)} l$, and $\partial^{(1,0)} l$ are linearly independent. We can see that all other partial derivatives $\partial^\nu l$, $1 \leq |\nu|$, are functions of $\Phi(\xi_i)$, $\phi(\xi_i)$, and terms in the Equations (B.5) and (B.6). Then $\tilde{Z}_i$ can be written as

$$\Big(\partial^{(0,1)}l_i,\ \partial^{(1,0)}l_i,\ \delta_i\phi(\xi_i),\ \delta_i\phi'(\xi_i),\ \delta_i\phi''(\xi_i),\ \delta_i\phi(\xi_i)\xi_i,\ \delta_i\phi'(\xi_i)\xi_i,\ \delta_i\phi''(\xi_i)\xi_i,\ \delta_i\phi'(\xi_i)\xi_i^2,$$
$$\delta_i\phi''(\xi_i)\xi_i^2,\ \delta_i\phi''(\xi_i)\xi_i^3,\ \delta_i\phi''(\xi_i)\xi_i^4,\ \delta_i\Big)$$

Because the expectations of $\xi_i^j\phi(\xi_i)^k\Phi(\xi_i)^m$, $j+k+m>0$, are finite over an open set containing the true parameters, the expectations of $\partial^\nu l$, $1\le|\nu|$, are finite over the same open set. Thus Condition (A2) holds.

For right censoring and a location-scale distribution with a likelihood function satisfying Conditions (A1) and (A2), it can be shown by using Equation (B.7) that $I(\theta_0)=D(\theta_0)$ . The calculation is straight forward, we omit the detail here. Note that $D$ is the variance-covariance matrix of score function so it is nonnegative definite. If the determinant of $D$ is 0, then

$$\frac{\partial l(X_i;(\sigma,x_p))}{\partial x_p}=c\frac{\partial l(X_i;(\sigma,x_p))}{\partial \sigma}, \tag{B.11}$$

for all possible values of $X_i$, where $c$ is a constant. From Equations (B.8), (B.9), and (B.10) we see that (B.11) is not true for the SEV, lognormal, and loglogistic distributions. Thus $D$ is positive definite. Thus Condition (A3) holds.

## Methods in the Simulation Study

This section describes some technical aspects for the different CI/CB procedures that we have studied in this paper. For more details, see the given references.

**Log LR method (LLR).** The distribution of $W$ is approximately $\chi_1^2$. Thus an approximate $100(1-\alpha)\%$ confidence interval can be calculated from $\min\{W^{-1}(\chi^2_{(1-\alpha,1)})\}$ and $\max\{W^{-1}(\chi^2_{(1-\alpha,1)})\}$, where $W^{-1}[\cdot]$ is the inverse mapping and $\chi^2_{(1-\alpha,1)}$ is the $1-\alpha$ quantile of $\chi^2$ distribution with 1 degree of freedom.

**Log LR Bartlett corrected method (LLRB).** Let $W_B=W/\mathrm{E}(W)$. In general one must substitute an estimate for $\mathrm{E}(W)$ computed from one's data. For complicated problems (e.g., those involving censoring) it is necessary to estimate of $\mathrm{E}(W)$ by using simulation. Then an approximate $100(1-\alpha)\%$ confidence interval can be obtained by using $\min\{W_B^{-1}[\chi^2_{(1-\alpha,1)}]\}$ and $\max\{W_B^{-1}[\chi^2_{(1-\alpha,1)}]\}$.

**Modified signed root log LR method (MSRLLR).** Barndorff-Nielsen and Cox (1994, pp 201-206) proved that under some regularity conditions, $R_M$ can be approximated by normal distribution with error rate $O(1/n^{3/2})$ when there is no censoring. Let $l = \log L(\theta)$ and $\theta = (\psi, \lambda)$, where $\psi$ is a scalar parameter. The MSRLLR is defined as

$$R_M(\psi) = R(\psi) + \frac{1}{R(\psi)} \log \left[\frac{U(\psi)}{R(\psi)}\right],$$

where

$$U(\psi) = |l_{;\hat{\theta}}(\hat{\theta}) - l_{;\hat{\theta}}(\hat{\theta}_\psi),\ l_{\lambda;\hat{\theta}}(\hat{\theta}_\psi)|/\{|\mathcal{J}_{\lambda\lambda}(\hat{\theta}_\psi)||\mathcal{J}(\hat{\theta})|\},$$

$$l(\theta) = l(\theta;\hat{\theta},a),\ l_{;\hat{\theta}}(\hat{\theta}) = \left.\frac{\partial l(\theta;\hat{\theta},a)}{\partial\hat{\theta}}\right|_{(\hat{\theta};\hat{\theta},a)},\ l_{\lambda;\hat{\theta}}(\hat{\theta}_\psi) = \left.\frac{\partial l(\theta;\hat{\theta},a)}{\partial\lambda\partial\hat{\theta}}\right|_{(\hat{\theta}_\psi;\hat{\theta},a)}.$$

$R$ is the signed square root log likelihood ratio statistic, $a$ is an ancillary statistic, $\mathcal{J}$ is the local information matrix of $\theta$, and $\mathcal{J}_{\lambda\lambda}$ is the local information matrix of $\lambda$. Let $z_\alpha$ be the $\alpha$ quantile of normal distribution. The $100(1-\alpha)\%$ confidence limits can be obtained from $(R_M)^{-1}(\pm z_{\alpha/2})$.

**Parametric transformed bootstrap-$t$ method (PTBT).** Let g be a smooth monotone function generally chosen such that $g(\hat{\theta}_1)$ has range on whole real line. Let $\hat{\theta}_1$ be the ML estimator of $\theta_1$ and let $\hat{\theta}_1^*$ be the bootstrap version of the ML estimator. Let $z_{g(\hat{\theta}_1^*)_{(\alpha)}}$ be the $\alpha$ quantile of the distribution of $[g(\hat{\theta}_1^*) - g(\hat{\theta}_1)]/\widehat{\mathrm{se}}^*[g(\hat{\theta}_1)]$, where $\widehat{\mathrm{se}}^*[g(\hat{\theta}_1)]$ is the bootstrap version of $\widehat{\mathrm{se}}[g(\hat{\theta}_1)]$. We choose $\widehat{\mathrm{se}}[g(\hat{\theta}_1)]$ to be $g'(\hat{\theta}_1)\hat{I}_{\hat{\theta}}^{(1,1)}g'(\hat{\theta}_1)$, where $\hat{I}_{\hat{\theta}}$ is the local estimate of $I_{\theta}$. For estimating quantiles of a positive random variable we take $g$ to be the log transformation. An approximate $100(1-\alpha)\%$ confidence interval for $\theta_1$ can be computed from $g^{-1}\{g(\hat{\theta}_1) - z_{g(\hat{\theta}_1^*)_{(1-\alpha/2)}}\widehat{\mathrm{se}}[g(\hat{\theta}_1)]\}$ and $g^{-1}\{g(\hat{\theta}_1) - z_{g(\hat{\theta}_1^*)_{(\alpha/2)}}\widehat{\mathrm{se}}[g(\hat{\theta}_1)]\}$.

**Parametric bootstrap bias-corrected accelerated method (PBBCA).** Efron and Tibshirani (1993, Section 14.3) showed an easy way to obtain BCA confidence intervals. An approximate $100(1-\alpha)\%$ confidence interval is given by $(\hat{\theta}^*_{1(\alpha_1)}, \hat{\theta}^*_{1(\alpha_2)})$. Where $\hat{\theta}^*_{1(\alpha)}$ is the $\alpha$ quantile of the distribution of $\hat{\theta}_1^*$ and

$$\alpha_1 = \Phi\left(\hat{z}_0 + \frac{\hat{z}_0 + z_{\alpha/2}}{1 - \hat{a}(\hat{z}_0 + z_{\alpha/2})}\right),\quad \alpha_2 = \Phi\left(\hat{z}_0 + \frac{\hat{z}_0 + z_{1-\alpha/2}}{1 - \hat{a}(\hat{z}_0 + z_{1-\alpha/2})}\right),$$

$$\hat{z}_0 = \Phi^{-1}\left(\frac{\#\{\hat{\theta}_1^*(b) < \hat{\theta}_1\}}{B}\right),\quad \hat{a} = \frac{\sum_{i=1}^n (\hat{\theta}_{1[\cdot]} - \hat{\theta}_{1[i]})^3}{6\left[\sum_{i=1}^n (\hat{\theta}_{1[\cdot]} - \hat{\theta}_{1[i]})^2\right]^{3/2}}.$$

Usually $\Phi$ is taken to be the standard normal cdf. Here $\hat{\theta}_{1[i]} = \hat{\theta}_1(X_{[i]})$, $X_{[i]}$ is the original sample with the $i$th point $x_i$ deleted, $\hat{\theta}_{1[\cdot]} = \sum_{i=1}^n \hat{\theta}_{1[i]}/n$, $z_\alpha$ is the $\alpha$ quantile of normal distribution, and $B$ is the number of the bootstrap samples, and $\widehat{\theta}_1^*(b), b = 1, \ldots, B$ are bootstrap versions of $\widehat{\theta}_1$.

If there is an increasing function $v_n$ (the exact form need not be known) such that

$$\Pr\left\{\frac{v_n(\widehat{\theta}_1) - v_n(\theta_1)}{1 + a v_n(\theta_1)} + z_0 \leq x\right\} = \Phi(x),$$

then the $\mathrm{BC}_a$ CI procedure is exact.

**Parametric bootstrap signed square root LLR method (PBSRLLR).** Suppose that $r_{\widehat{\theta}^*_{1(\alpha)}}$ is the $\alpha$ quantile of the bootstrap distribution of a SRLLR statistic, $R(\theta_1)$. Then an approximate $100(1-\alpha)\%$ confidence interval can be computed from $\min\{R^{-1}(r_{\widehat{\theta}^*_{1(\alpha/2)}}), R^{-1}(r_{\widehat{\theta}^*_{1(1-\alpha/2)}})\}$ and $\max\{R^{-1}(r_{\widehat{\theta}^*_{1(\alpha/2)}}), R^{-1}(r_{\widehat{\theta}^*_{1(1-\alpha/2)}})\}$.

**Parametric bootstrap modified signed square root LLR method (PBMSRLLR)** Let $R^*_M$ be the bootstrap version of $R_M$. Suppose $r_{\widehat{\theta}^*_{1(\alpha)}}$ be the $\alpha$ quantile of the distribution of $R^*_M$.The $100(1-\alpha)\%$ confidence limits are $\min\{{R_M}^{-1}(r_{\widehat{\theta}^*_{1(1-\alpha/2)}}), {R_M}^{-1}(r_{\widehat{\theta}^*_{1(\alpha/2)}})\}$ and $\max\{{R_M}^{-1}(r_{\widehat{\theta}^*_{1(1-\alpha/2)}}), {R_M}^{-1}(r_{\widehat{\theta}^*_{1(\alpha/2)}})\}$.

# APPENDIX C RESULTS USED FOR CONSTRUCTING SIMULTANEOUS BANDS

## Two-sided Simultaneous Confidence Bands

**Result 1.** In a location-scale model, the lower and upper confidence curves for quantiles are the same as the upper and lower confidence curves for the cdf, if those curves are computed from a convex joint confidence region.

**Proof.** We want to show that $[\widehat{x}_p(\min), p]$ and $[\widehat{x}_p(\max), p]$, $0 < p < 1$, are two curves the same as $[x, \max_{(\mu,\sigma)\in\mathcal{R}} \Phi((x-\mu)/\sigma)]$ and$[x, \min_{(\mu,\sigma)\in\mathcal{R}} \Phi((x-\mu)/\sigma)]$. We only show the lower confidence curve case, the upper case can be obtained analogously. Given $x$ on real line, there is a $p$ such that $\widehat{x}_p(\max) = x$. The lower confidence curve for $\Phi$ is $[x, \min_{(\mu,\sigma)\in R} \Phi((x-\mu)/\sigma)]$. The claim will be established if we show that $\min_{(\mu,\sigma)\in R} \Phi((x-\mu)/\sigma)$ is equal to $p$. That is $(\widehat{x}_p(\max), p) = [x, \min_{(\mu,\sigma)\in R} \Phi((x-\mu)/\sigma)]$, $0 < p < 1$. Suppose $\min_{(\mu,\sigma)\in R} \Phi((x-\mu)/\sigma)$ equals to $p_0$. Clearly $p_0 \leq p$ and there is at least a point $(\mu_{p_0}, \sigma_{p_0})$ in $\mathcal{R}$ satisfying equation (4.6). Suppose that $p_0$ is smaller than $p$. Then it follows that $-u_{p_0}^{-1}$ is also smaller than $-u_p^{-1}$. This means the line that passes thought the point $(\mu_{p_0}, \sigma_{p_0})$ (inside $\mathcal{R}$) with intercept $\widehat{x}_p(\max)$ is on the right of the tangent line of the region $\mathcal{R}$ with the same intercept $\widehat{x}_p(\max)$ (see Figure 4.1 for visual justification). This is impossible. So we have that $p_0 = p$.

**Result 2.** In a location-scale model, a two-sided simultaneous confidence band $\mathcal{B}$ has the same confidence level as its corresponding convex confidence region $\mathcal{R}$.

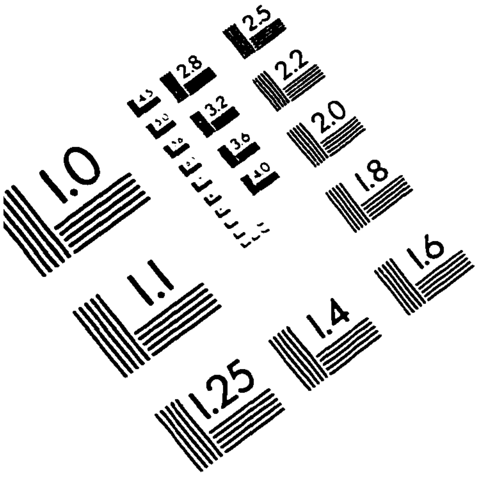
**Proof.** We consider any point $(\mu_0, \sigma_0)$ which is not in the region $\mathcal{R}$. Clearly there is a $p$ such that the line with slope $-u_p$ passes the point $(\mu_0, \sigma_0)$ but does not cross the region $\mathcal{R}$ (see Figure 4.2 for visual justification). This implies that the point $(\mu_0 + u_p\sigma_0, p)$ is not located in the band $\mathcal{B}$. So we conclude that no other points outside the region $\mathcal{R}$ will produce a cdf which lies entirely in the band $\mathcal{B}$. The band $\mathcal{B}$ hence has the same confidence level as the confidence region $\mathcal{R}$.

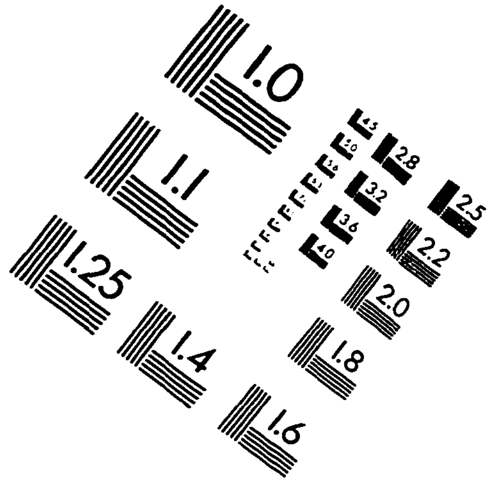
## One-sided Simultaneous Confidence Bands

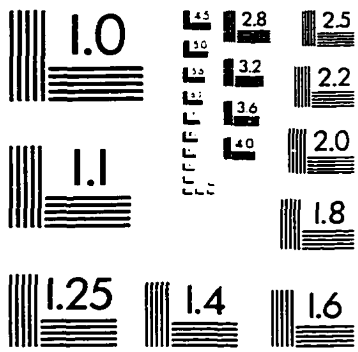
**Result 3.** In a location-scale model, the confidence level of one-sided confidence band of the cdf is the same as the confidence level of its corresponding convex confidence region $\mathcal{R}_l$.

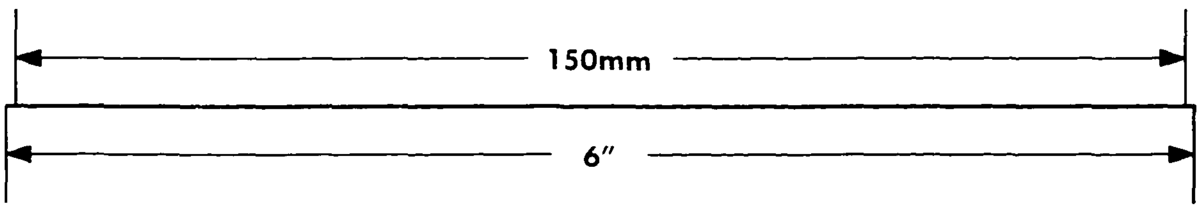
**Proof.** We only show the lower confidence band case, the upper case can be obtained analogously. If $(\mu_0, \sigma_0)$ is not in the region $\mathcal{R}_l$, there is at least a $p_0$ such that the line $\mu + u_{p_0}\sigma = x_{p_0}$ passing through the point $(\mu_0, \sigma_0)$ does not cross the region $\mathcal{R}_l$. Then the number $\mu_0 + u_{p_0}\sigma_0$ is bigger than $\widehat{x}_{p_0}(\max)$. This implies no other points outside region $\mathcal{R}_l$ could produce a confidence curve which lies entirely in the band $\mathcal{B}_l = \{(\widehat{x}_p(\max), p) : 0 < p < 1\}$ So the confidence level of $\mathcal{B}_l$ is the same as that of $\mathcal{R}_l$. That is the one-sided simultaneous confidence band will be exact if the corresponding convex confidence region $\mathcal{R}_l$ posses requested confidence level.

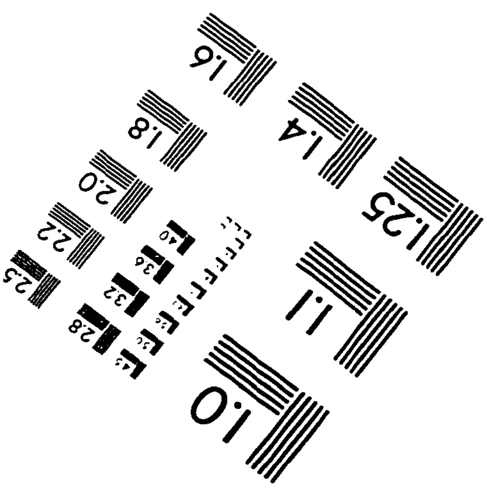
# IMAGE EVALUATION TEST TARGET (QA-3)

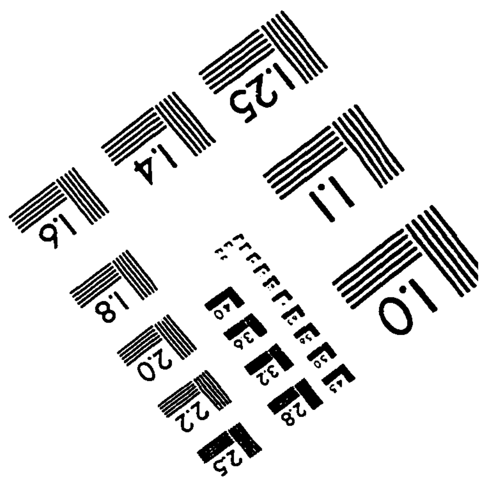




APPLIED IMAGE, Inc
1653 East Main Street
Rochester, NY 14609 USA
Phone: 716/482-0300
Fax: 716/288-5989